AMERICAN JOURNAL OF MATHEMATICS

FOUNDED BY THE JOHNS HOPKINS UNIVERSITY

PUBLISHED UNDER THE JOINT AUSPICES OF

THE JOHNS HOPKINS UNIVERSITY
AND
THE AMERICAN MATHEMATICAL SOCIETY

VOLUME LXXIV

1952

THE JOHNS HOPKINS PRESS
BALTIMORE 18, MARYLAND
U. S. A.

PRINTED IN THE UNITED STATES OF AMERICA
BY J. H. FURST COMPANY, BALTIMORE, MARYLAND

INDEX

ON THE PICARD VARIETIES ATTACHED TO ALGEBRAIC VARIETIES.*

By JUN-ICHI IGUSA.

In spite of the numerous classical investigations the theory of Picard varieties is not so complete as the special case of Jacobian varieties. In order to illustrate this circumstance, we should like to mention the problem of the relation between the abelian variety attached to the period matrix of the Picard integrals of the first kind on the given algebraic variety and its Picard variety. We could not find any treatment of such a fundamental problem in the literature.[1]

Now in modern algebraic geometry we have the powerful tool of "harmonic integrals," which has been invented by de Rham and Hodge and fully investigated by Kodaira and de Rham recently, together with the rigorous foundation of algebraic geometry due to van der Waerden, Chow, Zariski and Weil. On the base of these achievements we shall develop a theory of Picard varieties which includes the complete solution of the above problem.

In § I, we shall prove (for the sake of completeness), that continuous equivalence in the sense of Severi implies homological equivalence over integers. In § II, we shall define the Albanese variety A attached to a given variety V and also a skew-symmetric integral matrix E. In § III, we shall define a continuous family of divisors on V, which is "parametrized" by A, after Poincaré's method. The Picard variety $\mathfrak{P}$ attached to V is then defined by a certain "involution" of A, which can be described by the above matrix E. Our results include Weil's duality between abelian varieties. This paper contains two Appendices; in the second we shall sketch a correspondence theory between algebraic varieties.

While the outline of this paper was made in Tokyo in August 1949, I have received in writing it down at Kyoto constant encouragement and kind advice from Prof. Akizuki and valuable suggestions from Prof. Weil, to whom I wish to express my deepest appreciation.

* Received July 12, 1950; revised June 13, 1951.

[1] Indeed these two abelian varieties have been confused by some authors perhaps from the fact that they coincide if the given variety is a curve. On the other hand I heard from Weil by his kind letter of October 19, 1949 that he had also remarked the distinction between the abelian variety and its Picard variety; see § III, 10 of this paper.

§ I. Continuous Families and Continuous Equivalences on a Projective Model V.

1. Let L^n be a projective space in the algebraic geometry with the universal domain K of all complex numbers.[2] Let H^{n-r} be a linear variety in L^n, defined by a set of equations

$$\sum_{j=0}^{n} u_{ij}X_j = 0 \qquad (1 \leqq i \leqq r),$$

where $(u) = (u_{10}, u_{11}, \cdots, u_{rn})$ is a set of $r(n+1)$ independent variables over a field k. Such a linear variety was systematically used for the first time by van der Waerden; we shall call it a *generic linear variety over k*.

Now let V^d be any compact complex analytic variety of (complex) dimension d in L^n; by a recent result of Chow,[3] V is then an algebraic variety. We shall assume that V has no multiple point. Although we shall consider the case $d \geqq 2$, our results hold trivially in the case $d = 1$. In the following we shall assume that every field contains the smallest field of definition of V.

We shall start from the following lemma, which is slightly different from a similar lemma of Zariski.[4]

LEMMA. *Let L be a regular extension of a field K, and let ξ, η be two independent variables in L over K. If u is a variable over L, then $L(u)$ is regular over the field $K(u)(\xi + u\eta)$.*

Proof. As is readily seen, it is sufficient to consider the case where L is an algebraic extension over $K(\xi, \eta)$. Let u_i $(1 \leqq i \leqq N)$ be a set of N independent variables over L, and let K_i be the algebraic closure of $K(u_1, \cdots, u_N)(\xi + u_i\eta)$ in $L(u_1, \cdots, u_N)$. Then, since

$$[L(u_1, \cdots, u_N) : K(u_1, \cdots, u_N)(\xi, \eta)] = [L : K(\xi, \eta)],$$

if we take N sufficiently large, there must exist at least one pair (i, j) $(i \neq j)$ such that $K_i(\xi + u_j\eta) = K_j(\xi + u_i\eta)$. On the other hand since K_j is regular over $K(u_1, \cdots, u_N)$, $K_j(\xi + u_i\eta)$ is regular over $K(u_1, \cdots, u_N)(\xi + u_i\eta)$.

---

[2] We shall use the results and terminology of Weil's book: *Foundations of algebraic geometry*, American Mathematical Society Colloquium Publication no. 29 (1946). We shall cite this book as (F).

[3] W. L. Chow, "On compact complex analytic varieties," this JOURNAL, Vol. 71 (1949); theorem 5. We shall cite this paper as (C).

[4] O. Zariski, "Pencils on an algebraic variety and a new proof of a theorem of Bertini," *Transactions of the American Mathematical Society*, vol. 50 (1941); lemma 5.

Therefore $K_i = K(u_1, \cdots, u_N)(\xi + u_i\eta)$, and $L(u_1, \cdots, u_N)$ is regular over $K(u_1, \cdots, u_N)(\xi + u_i\eta)$; hence $L(u_i)$ is regular over $K(u_i)(\xi + u_i\eta)$.

Let $(x) = (x_0, x_1, \cdots, x_n)$ be a representative of a generic point of V over a field k, and say $x_0 = 1$; let $(u_1, \cdots, u_n)$ be a set of independent variables over $k(x)$, and put $u_0 = -\sum_{i=1}^{n} u_i x_i$. Then the equation $\sum_{j=0}^{n} u_i X_i = 0$ defines a generic linear variety H^{n-1} in L^n over k. Concerning the section of V by such an H, we have the following basic result.

PROPOSITION 1. *The intersection-product* $\mathsf{V} \cdot \mathsf{H}$ *is an (absolutely irreducible) variety without multiple point.*

Proof. Let x_{i_0} be a variable among the x_i $(1 \leqq i \leqq n)$ over k, and put

$$\xi = \sum_{0 \neq i \neq i_0} u_i x_i, \qquad \eta = x_{i_0};$$

then ξ, η are independent variables over $k(u_1, \cdots, u_{i_0-1}, u_{i_0+1}, \cdots, u_n)$, hence by the above lemma, $k(u_1, \cdots, u_n, x)$ is regular over $k(u_0, u_1, \cdots, u_n)$. Therefore the point of V with the representative (x) has a locus W over $k(u_0, u_1, \cdots, u_n)$ such that $\mathsf{V} \cdot \mathsf{H} = i(\mathsf{V} \cdot \mathsf{H}, \mathsf{W})\mathsf{W}$. On the other hand, H is transversal to V along W, hence by the criterion of multiplicity 1, we have $i(\mathsf{V} \cdot \mathsf{H}, \mathsf{W}) = 1$. Moreover by the "theorem of Bertini" on the variable singular points, W has no multiple point.

By repeated use of this proposition we conclude that the intersection-product W^{d-r} $(0 \leqq r < d)$ of V^d and a generic linear variety H^{n-r} over a field k is a subvariety of V without multiple points; we shall call it a *generic* $(d-r)$*-section of* V^d *over* k.

2. Now let Y be a positive r-cycle $(0 \leqq r \leqq d)$ on V and let M be the coefficients of the "associated form" of Y. With Chow we shall consider M as a set of homogeneous coordinates of a point in a projective space; we shall call it the *Chow point* of Y. We shall also identify M with the point itself. By an elementary property of the Chow point,[5] if M' is a specialization of M over a field k, there exists a rational r-cycle Y' with the Chow point M' over $k(\mathsf{M}')$. It is natural to call Y' a *specialization of* Y *over* k. Furthermore let Y be an r-cycle on V and let $\mathsf{Y} = \mathsf{Y}_1 - \mathsf{Y}_2$ be the reduced expression of Y as a difference of positive cycles Y_1 and Y_2 then if Y'_1 and Y'_2 are the specializations of Y_1 and Y_2 over k, we say that $\mathsf{Y}' = \mathsf{Y}'_1 - \mathsf{Y}'_2$ is a specializa-

---

[5] For a systematic treatment of the Chow point see v. d. Waerden's book: *Einführung in die algebraische Geometrie,* Berlin (1939), §§ 36-38.

tion of Y over k. On the other hand let K be an extension of k over which both Y and Y' are rational, and let H^{n-r} be a generic linear variety over K. It can be readily seen that Y' is a specialization of Y over k if and only if the 0-cycle $\mathsf{Y}' : \mathsf{H}$ is a specialization of the 0-cycle $\mathsf{Y} \cdot \mathsf{H}$ over $k(u)$, where (u) denotes the set of coefficients of the equations for H. Since our specialization of cycles is based on the point specialization, the basic properties of the usual specializations are valid for this general specialization; we shall show

PROPOSITION 2. *Let* U *be an abstract variety defined over a field* k; *let* M *be a generic point of* U *over* k *and let* M' *be a simple point of* U. *Let* X *be a rational* $(\mathsf{U} \times \mathsf{V})$*-cycle over* k *such that the* V*-cycle*

$$\mathsf{X}(\mathsf{M}') = \mathrm{pr}_{\mathsf{V}}\,(\mathsf{X} \cdot (\mathsf{M}' \times \mathsf{V}))$$

is defined. Then the similar cycle $\mathsf{X}(\mathsf{M})$ *has a uniquely determined specialization* $\mathsf{X}(\mathsf{M}')$ *over* $\mathsf{M} \to \mathsf{M}'$ *with reference to* k.

Proof. Let Y^r be a specialization of $\mathsf{X}(\mathsf{M})$ over $\mathsf{M} \to \mathsf{M}'$ with reference to k, and let K be an extension of $k(\mathsf{M}, \mathsf{M}')$ over which Y is rational; let H^{n-r} be generic linear variety over K. Then $\mathsf{Y} \cdot \mathsf{H}$ is a specialization of $\mathsf{X}(\mathsf{M}) \cdot \mathsf{H}$ over $\mathsf{M} \to \mathsf{M}'$ with reference to $k(u)$. On the other hand let W be the generic $(d-r)$-section of V by H over k, then we have

$$\mathsf{X}(\mathsf{M}) \cdot \mathsf{H} = \mathrm{pr}_{\mathsf{W}}\,((\mathsf{X}_0 \cdot (\mathsf{M}' \times \mathsf{W}))_{\mathsf{U}\times\mathsf{W}}),$$

$$\mathsf{X}(\mathsf{M}') \cdot \mathsf{H} = \mathrm{pr}_{\mathsf{W}}\,((\mathsf{X}_0 \cdot (\mathsf{M} \times \mathsf{W}))_{\mathsf{U}\times\mathsf{W}}),$$

where $\mathsf{X}_0 = \mathsf{X} \cdot (\mathsf{U} \times \mathsf{W})$. Therefore by Th. 13 of (F), Chap. VII, § 6, the W-cycle $\mathsf{X}(\mathsf{M}) \cdot \mathsf{H}$ has the *uniquely determined* specialization $\mathsf{X}(\mathsf{M}') \cdot \mathsf{H}$ over $\mathsf{M} \to \mathsf{M}'$ with reference to $k(u)$, hence we must have $\mathsf{Y} \cdot \mathsf{H} = \mathsf{X}(\mathsf{M}') \cdot \mathsf{H}$. However since Y and $\mathsf{X}(\mathsf{M}')$ are rational over K, and since H is a generic linear variety over K, this implies $\mathsf{Y} = \mathsf{X}(\mathsf{M}')$, which completes our proof.

Now let U, X, k and M be the same as in Proposition 2. but U be in the following a complete variety; then the specialization of $\mathsf{X}(\mathsf{M})$ over k constitute what we call the *continuous family determined by* X *on* V, or shortly the *continuous family* X. It can be readily verified that this definition does not depend on the choice of k and M, when X is given. Moreover we can and shall assume that every component of X has the projection U on U. We say that two r-cycles Y and Y' on V are *continuously equivalent*, if there exists a series of r-cycles $\mathsf{Y} = \mathsf{Y}_1, \mathsf{Y}_2, \cdots, \mathsf{Y}_{N+1} = \mathsf{Y}'$, a series of complete non-singular curves $\Gamma_1, \Gamma_2, \cdots, \Gamma_N$ and a series of $(\Gamma_i \times \mathsf{V})$-cycles X_i, such that

$$\mathsf{Y}_i = \mathrm{pr}_{\mathsf{V}}\,(\mathsf{X}_i \cdot (\mathsf{M}_i \times \mathsf{V})), \qquad \mathsf{Y}_{i+1} = \mathrm{pr}_{\mathsf{V}}\,(\mathsf{X} \cdot (\mathsf{M}'_i \times \mathsf{V})),$$

where M_i and M'_i are the points of Γ_i for $1 \leqq i \leqq N$. The continuous equivalence is thus an equivalence relation in the set of all r-cycles on V. Moreover since the points M_i and M'_i are homologous on Γ_i, the products $M_i \times V$ and $M'_i \times V$ are homologous on $\Gamma_i \times V$. Therefore the intersection products $X_i \cdot (M_i \times V)$ and $X_i \cdot (M'_i \times V)$ are homologous on $\Gamma_i \times V$, and their algebraic projections Y_i and Y_{i+1} on V are homologous on V. It follows, finally that

$$Y - Y' = \sum_{i=1}^{N} (Y_i - Y_{i+1})$$

is homologous to 0 on V. Here the homologies are taken over Z;[6] and all these follow from the topological intersection-theory.[7]

3. Now we shall connect our two notions by the following assertion.

THEOREM 1. *Any two cycles in the same continuous family on* V *are continuously equivalent.*

Proof. We shall use the same notations as before. Let Y be any specialization of $X(M)$ over k; we have only to show that they are continuously equivalent. Since U is a complete variety, there exists a point M' on U such that Y is a specialization of $X(M)$ over $M \rightarrow M'$ with reference to k. We shall first assume that U is a curve and let Γ be derived from U by normalization with reference to k; then Γ is a complete non-singular curve, which corresponds to U by a birational correspondence T over k. Let $M \times M^*$ be the generic point of T over k with the projection M on U; and let $M^{*\prime}$ be a specialization of M^* over the specialization $X(M) \rightarrow Y$, $M \rightarrow M'$ with reference to k. If we interchange the second and the third factors of the product $U \times \Gamma \times V \times V$, its subvariety $T \times \Delta$ corresponds biregularly to a birational correspondence Z between $U \times V$ and $\Gamma \times V$. Since Z is biregular along every component of X, the intersection-product $Z \cdot (X \times \Gamma \times V)$ is defined; we put $X^* = \mathrm{pr}_{\Gamma \times V}(Z \cdot (X \times \Gamma \times V))$. Then it can be readily verified that we have

$$X^*(M^*) = \mathrm{pr}_V (X^* \cdot (M^* \times V)) = X(M).$$

Moreover since every component of X^* has the projection Γ on Γ, the intersection-product $X^* \cdot (M^{*\prime} \times V)$ is defined. Therefore by the previous

---

[6] As in Bourbaki we shall denote by Z the ring of rational integers and by Q its field of quotients.

[7] Cf. S. Lefschetz, *Topology*, American Mathematical Society Colloquium Publication no. 12 (1930).

proposition, its algebraic projection $\mathsf{X}^*(\mathsf{M}^{*\prime})$ on V is the *uniquely determined* specialization of $\mathsf{X}(\mathsf{M})$ over $\mathsf{M}^* \to \mathsf{M}^{*\prime}$ with reference to k. Since Y is also the specialization of $\mathsf{X}(\mathsf{M})$ over the same specialization, we must have $\mathsf{Y} = \mathsf{X}^*(\mathsf{M}^{*\prime})$. Hence Y and $\mathsf{X}(\mathsf{M})$ are continuously equivalent.

Now we shall prove that $\mathsf{X}(\mathsf{M})$ and $\mathsf{X}(\mathsf{M}')$ are continuously equivalent, whenever M and M' are generic points of U^s over k. Since this is true for $s = 1$, we can use induction on s assuming $s \geqq 2$. By taking another generic point of U over $k(\mathsf{M}, \mathsf{M}')$, if necessary, we may assume that M and M' are already independent over k. Let U, (x) and (x') be a set of representatives of U, M and M' in the same ambient space S^m; let $(u^\alpha) = (u_1^\alpha, \cdots, u_m^\alpha)$ $(\alpha = 0, 1)$ be two sets of $2m$ independent variables over $k(x, x')$ and put

$$v^0 = -\sum_{i=1}^{m} u_i^0 x_i, \qquad v^1 = -\sum_{i=1}^{m} u_i^1 x'_i.$$

Then the equations $\sum_{i=1}^{m} u_i^\alpha X_i + v^\alpha = 0$ define two linear varieties H_α^{m-1} in S^m such that the intersection-products $W_\alpha = U \cdot H_\alpha$ are (absolutely irreducible) varieties. Moreover the intersection-product $W = U \cdot H_0 \cdot H_1$ is, at least, a prime rational U-cycle over $k(u^0, v^0, u^1, v^1)$; and a generic point N of W over this field is a generic point of W_α over $k(u^\alpha, v^\alpha)$ for $\alpha = 0, 1$. Now let W_α^{s-1} and N be the subvarieties of U^s, which have the representatives W_α and N in U; since the components of $\mathsf{X} \cap (\mathsf{W}_\alpha \times \mathsf{V})$, which have the projection W_α on U, are all proper, we can form a $(\mathsf{W}_\alpha \times \mathsf{V})$-cycle X_α by such components. Then we have

$$\mathsf{X}_0(\mathsf{M}) = \mathrm{pr}_{\mathsf{V}}\left((\mathsf{X}_0 \cdot (\mathsf{M} \times \mathsf{V}))_{\mathsf{W}_0 \times \mathsf{V}}\right) = \mathsf{X}(\mathsf{M}),$$

$$\mathsf{X}_0(\mathsf{N}) = \mathsf{X}(\mathsf{N}) = \mathsf{X}_1(\mathsf{N}), \qquad \mathsf{X}_1(\mathsf{M}') = \mathsf{X}(\mathsf{M}');$$

therefore by the induction assumption, $\mathsf{X}(\mathsf{M})$ and $\mathsf{X}(\mathsf{M}')$ are continuously equivalent.

Finally, even if M' is not a generic point of U over k, we may assume that Y is the uniquely determined specialization of $\mathsf{X}(\mathsf{M})$ over $\mathsf{M} \to \mathsf{M}'$ with reference to k. Otherwise let M^* and $\mathsf{M}^{*\prime}$ be the Chow points of $\mathsf{X}(\mathsf{M})$ and Y; since M^* is rational over $k(\mathsf{M})$, it has a locus U^* over k. Moreover, since $\mathsf{X}(\mathsf{M})$ is rational over $k(\mathsf{M}^*)$, there exists a rational $(\mathsf{U}^* \times \mathsf{V})$-cycle X^* over k, every component of which has the projection U^* on U^*, such that

$$\mathsf{X}^*(\mathsf{M}^*) = \mathrm{pr}_{\mathsf{V}}\ (\mathsf{X}^* \cdot (\mathsf{M}^* \times \mathsf{V})) = \mathsf{X}(\mathsf{M}).$$

Then we have only to replace U, X and M' by U^*, X^* and $\mathsf{M}^{*\prime}$. Now by Prop. 7 of (F), Appendix II, there exists a generic specialization N of M

over k and an extension K of k such that N has a locus U_0 of dimension 1 over K containing M'. As before let X_0 be the $(\mathsf{U}_0 \times \mathsf{V})$-cycle composed of those components of $\mathsf{X} \cap (\mathsf{U}_0 \times \mathsf{V})$, which have the projection U_0 on U; we then have

$$\mathsf{X}_0(\mathsf{N}) = \mathrm{pr}_{\mathsf{V}}\ ((\mathsf{X}_0 \cdot (\mathsf{N} \times \mathsf{V}))_{\mathsf{U}_0 \times \mathsf{V}}) = \mathsf{X}(\mathsf{N}).$$

By assumption, Y is also the specialization of $\mathsf{X}(\mathsf{N})$ over $\mathsf{N} \to \mathsf{M}'$ with reference to K. Therefore since U_0 is a curve, Y and $\mathsf{X}(\mathsf{N})$, and so, by our second step, Y and $\mathsf{X}(\mathsf{M})$ are continuously equivalent.

§ II. Continuous Family of Curves in V.

4. Let W^1 be a generic 1-section of V over a field k and let M be the Chow point of W^1. Since W^1 is defined over a purely transcendental extension of k, $k(\mathsf{M})$ is regular over k and M has a locus U over k. Moreover there exists a rational $(\mathsf{U} \times \mathsf{V})$-cycle C over k, every component of which has the projection U on U, such that

$$\mathsf{C}(\mathsf{M}) = \mathrm{pr}_{\mathsf{V}}\ (\mathsf{C} \cdot (\mathsf{M} \times \mathsf{V})) = \mathsf{W}^1.$$

Since W^1 is a variety, C must also be a variety; we shall study the properties of this continuous family in V with a "generic curve" $\mathsf{C}(\mathsf{M})$ over k.

A differential form or a differential Φ on V, which can be expressed in the form $\Phi = \Sigma F_{i_1 \ldots i_s}(z)\, dz_{i_1} \wedge \cdots \wedge dz_{i_s}$, where the coefficients $F_{i_1 \ldots i_s}(z)$ and the "uniformizing parameters" $z_1, \cdots, z_d$ are (algebraic) functions on V^d, will be called an *algebraic differential of degree s on* V. Here $\wedge$ denotes Grassmann multiplication. An algebraic differential of degree 1 on V will be called a Picard differential on V. It can be proved easily by the last corollary in (C) that in the above definition we have only to assume that the coefficients $F_{i_1 \ldots i_s}(z)$ are meromorphic in $z_1, \cdots, z_d$ for every choice of $z_1, \cdots, z_d$. If $F_{i_1 \ldots i_s}(z)$ are always regular analytic in $z_1, \cdots, z_d$, Φ is said to be *of the first kind*. It is then a complex harmonic form in the sense of Hodge on the compact manifold V with a "natural" Kählerian metric.[8]

THEOREM 2. *The set of linearly independent Picard differentials of the first kind on* V *remains linearly independent on each generic curve* $\mathsf{C}(\mathsf{M})$ *over* k.

---

[8] Cf. W. V. D. Hodge, *The theory and applications of harmonic integrals*, Cambridge University Press (1941); Chap. IV. We shall cite this book as (H). The theorem which is stated above is implicite in (H), but can be proved as follows. By the special nature of the Kählerian metric, the "dual form" of every algebraic differential of the first kind is exact. Therefore if Φ is such a form, $d\Phi$ and its dual form are both exact. This shows precisely that $d\Phi$ is harmonic. Since it is at the same time a derived form, we have $d\Phi = 0$.

We shall give the "analytic" proof of Theorem 2 at the end of this section, and at present, we shall derive some of its consequences. Let p be the genus of $\mathsf{C}(\mathsf{M})$; since the generic curves are the transforms of one another in the sense of (F), Chap. VII, § 2, p does not depend on M. Moreover by Theorem 2, we can find a base of the Picard differentials of the first kind on V

$$(\Phi) = (\Phi_1, \Phi_2, \cdots, \Phi_q) \qquad (p \geqq q \geqq 0);$$

the number q is called the *irregularity* of V.

We shall denote by $\mathfrak{H}^r(\mathsf{V}, Z)$ or $\mathfrak{H}^r(\mathsf{V}, Q)$ the r-dimensional homology group of V modulo Z or Q, and by $R^r(\mathsf{V})$ the rank of $\mathfrak{H}^r(\mathsf{V}, Q)$. Hodge[9] gave an elegant proof of the classical theorem that $R^1(\mathsf{V}) = 2q$. Now let $\gamma_1, \cdots, \gamma_{2q}, \delta_1, \cdots, \delta_t$ be a base of $\mathfrak{H}^1(\mathsf{V}, Z)$, where $(\gamma) = (\gamma_1, \gamma_2, \cdots, \gamma_{2q})$ is a base of the 1-dimensional Betti group $\mathfrak{B}(\mathsf{V})$ of V modulo Z and where $\delta_1, \delta_2, \cdots, \delta_t$ are torsion cycles. Although the torsion cycles are absent if V is a curve, they actually appear in the general case. It is clear that the period matrix $\omega = (\int \gamma_i \Phi_\alpha)$ of the "Picard integrals" $\int \Phi_\alpha$ $(1 \leqq \alpha \leqq q)$ along the 1-cycles γ_i $(1 \leqq i \leqq 2q)$, is determined by V up to an "isomorphism" $\omega \to \Lambda\omega L$, where Λ is a non-singular K-matrix of degree q, and L a unimodular Z-matrix of degree $2q$.

On the other hand, let $(\alpha) = (\alpha_1, \alpha_2, \cdots, \alpha_{2p})$ be a base of $\mathfrak{H}^1(\mathsf{C}(\mathsf{M}), Z)$. Then there exists a uniquely determined Z-matrix A of type $(2q, 2p)$ such that $(\alpha) \sim (\gamma)A$ (mod. Q). In this paper the skew-symmetric Z-matrix $E_{\mathsf{M}} = A\,{}^tI_\alpha{}^{-1t}A$ plays an important rôle,[10] where I_α means the unimodular Z-matrix of Kronecker indices $I(\alpha_i, \alpha_j; \mathsf{C}(\mathsf{M}))$ on $\mathsf{C}(\mathsf{M})$. It can be readily verified that E_{M} does not depend on the choice of the base (α), if the base (γ) and the generic curve $\mathsf{C}(\mathsf{M})$ over k are given. We shall see later that E_{M} does not depend on M.

5. Applying the celebrated theorem of Poincaré in the theory of "reducible integrals"[11] to the above circumstances, we can find a base (β) of $\mathfrak{H}^1(\mathsf{C}(\mathsf{M}), Q)$, which is composed of "invariant cycles" $\beta_i \sim \gamma_i$ (mod. Q)

---

[9] Cf. (H), § 49. See also his original proof: "Harmonic integrals associated with algebraic varieties," *Proceedings of the London Mathematical Society*, vol. 39 (1935).

[10] Let M be a matrix with any coefficients, then ${}^tM$ means the transposed matrix of M. Moreover if M is a K-matrix, we shall denote by $\overline{M}$ the image of M by the involutive automorphism of K over the real field.

[11] Cf. O. Zariski, *Algebraic surfaces*, Ergebnisse der Mathematik (1935), Chap. VII, § 6.

$(i \leqq 2q)$, and of "vanishing cycles" $\beta_i \sim 0$ (in V mod. Q) $(i > 2q)$, such that if we define a square Q-matrix B of degree $2p$ by

$$(\alpha) \sim (\beta)B \qquad (\text{on } \mathsf{C}(\mathsf{M}) \text{ mod. } Q),$$

$$B^t I_\alpha^{-1t} B = \begin{pmatrix} \boxed{E_{\mathsf{M}}} & 0 \\ 0 & \boxed{} \end{pmatrix}.$$

Moreover we can find a "complementary set" of Abelian integrals $\int \Phi_{q+1}, \cdots, \int \Phi_p$ on $\mathsf{C}(\mathsf{M})$ such that the period matrix Ω_β of $\int \Phi_\alpha$ $(1 \leqq \alpha \leqq p)$ along the 1-cycles β_i $(1 \leqq i \leqq 2p)$ splits in the following form:

$$\Omega_\beta = \begin{pmatrix} \boxed{\omega} & 0 \\ 0 & \boxed{} \end{pmatrix}.$$

It follows that ${}^tE_{\mathsf{M}} = -E_{\mathsf{M}}$ is a "principal matrix" of the "Riemann matrix" ω.

Now the $2q$ columns of ω generate a discrete subgroup $[\omega]$ of rank $2q$ in the complex vector space S^q, and therefore the factor group $S^q/[\omega]$ is a complex toroid of (complex) dimension q. Moreover, since ω is shown to be a Riemann matrix, we see that this complex toroid is mapped in a one-to-one way to a non-singular variety A^q in a suitable projective space by means of theta functions $\Theta(v) = \Theta(v_1, \cdots, v_q)$, belonging to $[\omega]$.[12] It then follows from (C) that A^q is an abelian variety in the sense of Weil[13]; following him we shall call A the *Albanese variety attached to* V. Since the period matrix ω is determined up to an isomorphism, A is determined up to an isomorphism in the sense of (V) by our projective model V. We shall see later that A is attached to V in a birationally invariant way.

6. Now we shall prove our Theorem 2. There exists, by (H), Chap. IV, a harmonic 2-form ν in L^n such that $\int_{\mathsf{D}} \nu = 1$, where D is any projective 1-space in L^n. ν induces a similar form on V, which we shall denote by the same letter; and we shall show that ν and the generic $(d-1)$-section $\mathsf{W} = \mathsf{V} \cdot \mathsf{H}$ of V^d are homologous in the sense of de Rham.

LEMMA. *Let Z be any topological 2-cycle in* V. *Then* $\int_Z \nu = I(Z, \mathsf{W}; \mathsf{V})$.

---

[12] Cf. S. Lefschetz, "On certain numerical invariants of algebraic varieties with applications to abelian verieties," *Transactions of the American Mathematical Society*, vol. 22 (1921).

[13] A. Weil, "Variétés abeliennes et courbes algébriques," *Actualité Scientifiques et Industrielles*, no. 1064 (1948). We cite this book as (V).

Proof. Since we have $\int_Z \nu = I(Z, \mathsf{H})$, it is sufficient to show that the equality $Z \cdot \mathsf{H} = (Z \cdot (\mathsf{V} \cdot \mathsf{H}))_{\mathsf{V}}$. Thereby we may assume that Z is at least "differentiable" in the neighborhoods of the points in $Z \cap \mathsf{H}$ and that Z is transversal to H at these points. Then our formula surely holds locally, hence in the large.[14]

Now we can prove the following remarkable fact, which was discovered by Lefschetz.[15]

PROPOSITION 3. *Let* $Z_1, \cdots, Z_R$ $(R = R^{2d-r}(\mathsf{V}))$ *be a base of* $\mathfrak{H}^{2d-r}(\mathsf{V}, Q)$ $(0 \leqq r \leqq d)$. *Then the r-cycles*

$$\Gamma_i = (Z_i \cdot \mathsf{W})_{\mathsf{V}} \qquad (1 \leqq i \leqq R)$$

on the generic r-section W *of* V^d *are independent as cycles in* V modulo Q.

Proof. By Hodge's existence theorem[16] there exists a set of harmonic r-forms $\phi_1, \cdots, \phi_R$ in V such that $\int_\Gamma \phi_i = I(Z_i, \Gamma; \mathsf{V})$ $(1 \leqq i \leqq R)$ for every topological r-cycle Γ. Consider the harmonic $2(d-r)$-form $\nu_{d-r} = \nu \wedge \cdots \wedge \nu$ $((d-r)$-factors$)$. Then by Theorem III and by Lemma C, § 42 in (H), we conclude that the harmonic $(2r-r)$-forms $\phi_i \wedge \nu_{d-r}$ $(1 \leqq i \leqq R)$ are linearly independent. By de Rham's theorem, however, $\int_Z \phi_i \wedge \nu_{d-r} = I(\Gamma_i, Z; \mathsf{V})$ $(1 \leqq i \leqq R)$ for every topological $(2d-r)$-cycle Z; our proposition follows from this fact.

COROLLARY. *Every topological r-cycle in* V^d $(0 \leqq r \leqq d)$ *is homologous to some cycle on the generic s-section* W^s *of* V $(s \geqq r)$ modulo Q.

Theorem 2 is now a simple consequence of this corollary in the case of $r = s = 1$. However a purely algebraic proof for this theorem is greatly to be desired.

§ III. Poincaré Family and Picard Variety Attached to V.

7. We shall start from the following q "Abelian sums"

$$\sum_{i=1}^{q} \int^{\mathsf{P}_i} \Phi_\alpha = v_\alpha \qquad (1 \leqq \alpha \leqq q),$$

---

[14] See loc. cit. 7.

[15] S. Lefshetz, *L'analysis situs et la géométrie algébrique*, Paris (1924), Chap. V.

[16] Cf. K. Kodaira, "Harmonic fields in Riemannian manifolds," *Annals of Mathematics*, vol. 50 (1949).

where $\mathsf{P}_1, \cdots, \mathsf{P}_q$ are q points in V. If we restrict these points to the generic curve $\mathsf{C}(\mathsf{M})$ over k, we get an analytic mapping from the product $\tilde{\mathsf{C}}(\mathsf{M}) = \mathsf{C}(\mathsf{M}) \times \cdots \times \mathsf{C}(\mathsf{M})$ (q-factors) into the complex toroid $S^q/[\omega]$, which we shall denote by $\boldsymbol{\omega}$, in a natural way.

PROPOSITION 4. *The image $\tilde{\boldsymbol{\omega}}$ of $\tilde{\mathsf{C}}(\mathsf{M})$ covers the whole of $\boldsymbol{\omega}$, and covers in general exactly $q!\epsilon_{\mathsf{M}}$-times; where ϵ_{M} denotes the "Pfaffian" of E_{M}.*

Proof. Let $\mathsf{P}_1, \cdots, \mathsf{P}_q$ be a set of q distinct points on $\mathsf{C}(\mathsf{M})$; we shall show that the Jacobian of our mapping does not vanish there identically. In fact let $x_1, \cdots, x_q$ be the uniformizing parameters of $\mathsf{C}(\mathsf{M})$ at $\mathsf{P}_1, \cdots, \mathsf{P}_q$ respectively; then in the contrary case we have $\det |\Phi_\alpha(x_\beta)| = 0$, identically in $x_1, \cdots, x_q$. This, however, would contradict the linear independence of $\Phi_1, \cdots, \Phi_q$ on $\mathsf{C}(\mathsf{M})$. Therefore if we regard $\tilde{\boldsymbol{\omega}}$ as a compact connected "covering variety" of $\boldsymbol{\omega}$, its projection on $\boldsymbol{\omega}$ contains an open set of $\boldsymbol{\omega}$. Since the projection $\mathfrak{F}$ of the points of $\tilde{\boldsymbol{\omega}}$ on $\boldsymbol{\omega}$, at which the Jacobian vanishes, forms an analytic bunch of $(q-1)$ (complex) dimensions (at most) in $\boldsymbol{\omega}$, its complement $\boldsymbol{\omega} - \mathfrak{F}$ is connected. Our proposition follows from this fact except for the exact value of the covering; and this crucial result can be proved by Wirtinger's method [17] as follows.

Let $m(\boldsymbol{\omega})$ and $m(\tilde{\boldsymbol{\omega}})$ be the "volumes" of $\boldsymbol{\omega}$ and $\tilde{\boldsymbol{\omega}}$ respectively; and put

$$\int^{\mathsf{P}} \Phi_\alpha = y_{2\alpha-1} + i y_{2\alpha} \qquad (1 \leqq \alpha \leqq q).$$

Then it can be readily seen that

$$m(\tilde{\boldsymbol{\omega}}) = \sum_{i<j} \operatorname{sgn}\binom{i_1 j_1 \cdots i_q j_q}{1 \cdots\cdots 2q} \prod_{h=1}^{q} \oint y_{i_h} dy_{j_h},$$

where $\oint$ means the "boundary integral" on the Riemann surface of $\mathsf{C}(\mathsf{M})$. Therefore we have

$$m(\tilde{\boldsymbol{\omega}}) = m(\boldsymbol{\omega}) \sum_{i<j} \operatorname{sgn}\binom{i_1 j_1 \cdots i_q j_q}{1 \cdots\cdots 2q} e_{i_1 j_1} \cdots e_{i_q j_q},$$

where e_{ij} $(1 \leqq i, j \leqq 2q)$ are the coefficients of E_{M}; hence $m(\tilde{\boldsymbol{\omega}}) : m(\boldsymbol{\omega}) = q!(\det |E_{\mathsf{M}}|)^{\frac{1}{2}}$, which completes the proof.

---

[17] W. Wirtinger, "Zur Theorie der $2n$-fach periodischen Funktionen," *Monatshefte für Mathematik*, Bd. 7 (1896).

Now if we introduce a function Φ on V by writing

$$\Phi(\mathsf{P}) = \Theta\left(\int^{\mathsf{P}} \Phi_1, \cdots, \int^{\mathsf{P}} \Phi_q\right),$$

the function $\tilde{\Phi}$ on $\tilde{\mathsf{V}} = \mathsf{V} \times \cdots \times \mathsf{V}$ (q-factors)

$$\tilde{\Phi}(\mathsf{P}_1, \cdots, \mathsf{P}_q) = \Theta\left(\sum_{i=1}^{q} \int^{\mathsf{P}_i} \Phi_1, \cdots, \sum_{i=1}^{q} \int^{\mathsf{P}_i} \Phi_q\right),$$

can be written in the form

$$\tilde{\Phi}(\mathsf{P}_1, \cdots, \mathsf{P}_q) = \sum_{i=1}^{q} \Phi(\mathsf{P}_i);$$

where the right side refers to the group variety A. The fact that Φ and hence also $\tilde{\Phi}$ are (algebraic) functions follows from (C) Theorem 7. If M is a generic point of U over a common field of definition K of C, Φ, A and of the composition function in A, the functions Φ and $\tilde{\Phi}$ induce the function ϕ and $\tilde{\phi}$ on $\mathsf{C}(\mathsf{M})$ and $\tilde{\mathsf{C}}(\mathsf{M})$ respectively, such that their graphs are related as follows:

$$\Gamma_\Phi \cdot (\mathsf{C}(\mathsf{M}) \times \mathsf{A}) = \Gamma_\phi, \qquad \Gamma_{\tilde{\Phi}} \cdot (\tilde{\mathsf{C}}(\mathsf{M}) \times \mathsf{A}) = \Gamma_{\tilde{\phi}}.$$

Hence, using Proposition 4, we can obtain the following result, which is well known if V is a curve.

THEOREM 3. *There exists a function Φ on V with values in the Albanese variety A^q attached to V. Moreover on each generic curve $\mathsf{C}(\mathsf{M})$ over K, Φ induces a function ϕ such that if $\mathsf{P}_1, \cdots, \mathsf{P}_q$ are q independent generic points of $\mathsf{C}(\mathsf{M})$ over $K(\mathsf{M})$, then the point $z = \sum_{i=1}^{q} \phi(\mathsf{P}_i)$, is a generic point of A over $K(\mathsf{M})$ and satisfies*

$$[K(\mathsf{M})(\mathsf{P}_1, \cdots, \mathsf{P}_q)_s : K(\mathsf{M})(z)] = \epsilon_\mathsf{M};$$

where $K(\mathsf{M})(\mathsf{P}_1, \cdots, \mathsf{P}_q)_s$ is the invariant subfield of $K(\mathsf{M})(\mathsf{P}_1, \cdots, \mathsf{P}_q)$ by $q!$ permutations of $\mathsf{P}_1, \cdots, \mathsf{P}_q$.

We conclude from this theorem that ϵ_M does not depend on the choice of $\mathsf{C}(\mathsf{M})$; hence we may write it as ϵ.

Now the point $z \times \mathsf{P}_1$, say, in the product $\mathsf{A} \times \mathsf{V}$ is rational over $K(\mathsf{M}, \mathsf{P}_1, \cdots, \mathsf{P}_q)$; hence $K(z, \mathsf{P}_1)$ is regular over K and $z \times \mathsf{P}_1$ has a locus X over K. Moreover the V-cycle $\mathsf{X}(z) = \mathrm{pr}_\mathsf{V}(\mathsf{X} \cdot (z \times \mathsf{V}))$ is prime rational over $K(z)$ and has a non-empty intersection with the generic curve $\mathsf{C}(\mathsf{M})$ over $K(z)$. Since P_1 is algebraic over $K(\mathsf{M}, z)$, $\mathsf{X}(z)$ is a prime rational V-divisor over $K(z)$; and the intersection-product $\mathsf{X}(z) \cdot \mathsf{C}(\mathsf{M})$ is a prime

rational $C(M)$-divisor over $K(M, z)$. On the other hand, the $C(M)$-divisor $\mathrm{pr}_{C(M)}(\mathfrak{F}^{-1}(z))$ consists of $q\epsilon$ conjugate points of P_1 over $K(M, z)$, each being repeated $(q-1)!$-times; therefore we have

$$X(z)\cdot C(M) = 1/(q-1)!\,\mathrm{pr}_{C(M)}(\mathfrak{F}^{-1}(z)).$$

For historical reasons we shall attach the name of *Poincaré family* to this continuous family X. If Y is a member of the Poincaré family and if z' is the specialization of z over the specialization $X(z)\to Y$ with reference to K, we shall write $Y = X(z')$. Since such $X(z')$ are continuously equivalent in the sense of § I, 2, they are homologous to each other modulo Z.

8. If a V-divisor Y is homologous to 0 modulo Q, there exists a Picard differential of the "third kind" Ψ on V with the "residue" Y. This is known as Lefschetz's theorem [18] and proved elegantly by Weil.[19] We may assume that the real part of the integral of Ψ is one-valued on V, for otherwise we have only to add a suitable linear combination of Φ_α $(1\leqq\alpha\leqq q)$ to Ψ. Now let δ be any torsion 1-cycle in V such that $m\delta\sim 0$ (mod. Z); let C be a 2-chain over Z with the boundary $m\delta$. Then it can be readily seen from Stokes theorem that

$$\int_\delta \Psi = 2\pi i/mI(C, Y; V) = 2\pi i L(\delta, Y; V),$$

where $L(\delta, Y; V)$ denotes the linking coefficient of δ and Y on V. Therefore if we put $F(P) = \exp(\int^{P}\Psi)$, F is meromorphic on V; and if we continue F analytically along a continuous closed path Γ in V, F is multiplied by a constant factor $\chi_Y(\Gamma)$ of absolute value 1, which depends on the homology class of Γ modulo Z only. Now by the preparation theorem of Weierstrass,[20] we can define the V-divisor (F) for such a "multiplicative function" F on V, and we have $(F) = Y$. Since the Picard integral of the first kind with pure imaginary periods must be a constant (as follows from the fact that ω is a Riemann matrix), F is uniquely determined up to a constant factor by Y. Therefore each Y determines *uniquely* a character $\chi_Y(\Gamma)$ of the discrete group $\mathfrak{H}^1(V, Z)$. For a torsion cycle δ, we have

$$\chi_Y(\delta) = \exp(2\pi i L(\delta, Y; V));$$

[18] Cf. loc. cit. 15, note I.

[19] A. Weil, "Sur la théorie des formes différentielles attaché à une variété analytique complexe," *Comment. Math. Helv.*, vol. 20 (1947).

[20] Cf. S. Bochner and W. T. Martin, *Several complex variables*, Princeton University Press (1948), Chap. IX.

hence $\chi_{\mathsf{Y}}(\delta)=1$, whenever Y is homologous to 0 modulo Z. Such a Y thus induces a character of the Betti group $\mathfrak{B}(\mathsf{V})$; and the difference $\mathsf{Y}=\mathsf{X}(z)-\mathsf{X}(z')$ is surely such a V-divisor. In this case we shall obtain an explicit formula of the character

$$(\chi_{\mathsf{Y}}(\gamma))=(\chi_{\mathsf{Y}}(\gamma_1),\cdots,\chi_{\mathsf{Y}}(\gamma_{2q})).$$

9. We may first assume, by taking a suitable extension of K if necesary, that $\mathsf{X}(z')$ is rational over K. Let L be an extension of K over which $\mathsf{X}(z)$ is rational and let M and w be independent generic points of U and A over L. Then the intersection-products $\mathsf{X}(z)\cdot\mathsf{C}(\mathsf{M})$ and $\mathsf{X}(w)\cdot\mathsf{C}(\mathsf{M})$ are defined, and by definition the former is a specialization of the latter over $w\to z$ with reference to $L(\mathsf{M})$. On the other hand we have

$$\mathsf{X}(w)\cdot\mathsf{C}(\mathsf{M})=\sum_{i=1}^{\epsilon}\sum_{j=1}^{q}\mathsf{Q}_{ij}$$

with $\sum_{j=1}^{q}\phi(\mathsf{Q}_{ij})=w\quad(1\leqq i\leqq\epsilon)$ by Theorem 3. Hence we may write $\mathsf{X}(z)\cdot\mathsf{C}(\mathsf{M})$ in the form

$$\mathsf{X}(z)\cdot\mathsf{C}(\mathsf{M})=\sum_{i=1}^{\epsilon}\sum_{j=1}^{q}\mathsf{P}_{ij},$$

where $(\cdots,\mathsf{P}_{ij},\cdots)$ is the specialization of $(\cdots,\mathsf{Q}_{ij},\cdots)$ over $w\to z$ with reference to $L(\mathsf{M})$. Since the (algebraic) function $w=\Phi(\mathsf{Q}_{i1},\cdots,\mathsf{Q}_{iq})$ is "defined" at $(\mathsf{P}_{i1},\cdots,\mathsf{P}_{iq})$, the "value" of Φ at this point is the *uniquely determined* specialization of w over $(\mathsf{Q}_{i1},\cdots,\mathsf{Q}_{iq})\to(\mathsf{P}_{i1},\cdots,\mathsf{P}_{iq})$ with reference to $L(\mathsf{M})$. We therefore have

$$z=\tilde{\Phi}(\mathsf{P}_{i1},\cdots,\mathsf{P}_{iq})=\Theta\Big(\sum_{j=1}^{q}\int^{\mathsf{P}_{ij}}\Phi_1,\cdots,\sum_{j=1}^{q}\int^{\mathsf{P}_{ij}}\Phi_q\Big)\qquad(1\leqq i\leqq\epsilon).$$

Similarly $\mathsf{X}(z')\cdot\mathsf{C}(\mathsf{M})$ can be written in the form

$$\mathsf{X}(z')\cdot\mathsf{C}(\mathsf{M})=\sum_{i=1}^{\epsilon}\sum_{j=1}^{q}\mathsf{P}'_{ij},$$

and we have

$$z'=\Theta\Big(\sum_{j=1}^{q}\int^{\mathsf{P}'_{ij}}\Phi_1,\cdots,\sum_{j=1}^{q}\int^{\mathsf{P}'_{ij}}\Phi_q\Big)\qquad(1\leqq i\leqq\epsilon).$$

Therefore if $z-z'$ has the "coordinates" (v), then

$$\sum_{i=1}^{\epsilon}\sum_{j=1}^{q}\int_{\mathsf{P}'_{ij}}^{\mathsf{P}_{ij}}(\Phi)=\epsilon\cdot(v),$$

where $(\Phi)=(\Phi_1,\cdots,\Phi_q)$.

Now the multiplicative function F on V induces a similar function f on $\mathsf{C}(\mathsf{M})$ such that

$$(f) = (F)\cdot \mathsf{C}(\mathsf{M}) = \mathsf{X}(z)\cdot \mathsf{C}(\mathsf{M}) - \mathsf{X}(z')\cdot \mathsf{C}(\mathsf{M}).$$

But since $(\gamma_1,\cdots,\gamma_{2Q},0,\cdots,0)\sim(\beta)\sim(\alpha)\cdot B^{-1}$ (mod. Q), it follows that if e is a natural number which makes eB^{-1} into a Z-matrix, and if $(\phi(\alpha)) = (\phi(\alpha_1),\phi(\alpha_2),\cdots,\phi(\alpha_{2p}))$ is a base of the harmonic 1-forms on $\mathsf{C}(\mathsf{M})$, which is homologous to (α) in the sense of de Rham, then we have

$$(\chi_{\mathsf{Y}}(\gamma)^e\chi_{\mathsf{Y}}(0)^e) = \exp(2\pi i \sum_{ij}\int_{\mathsf{P}'_{ij}}^{\mathsf{P}_{ij}}(\phi(\alpha))(eB^{-1})).^{21}$$

On the other hand there exists a K-matrix X of type $(p,2p)$ such that $(\phi(\alpha)) = (\Phi\bar{\Phi})\begin{pmatrix} X \\ \bar{X}\end{pmatrix}$, where (Φ) has p columns. By integrating this equation, we get $I_\alpha = {}^tB({}^t\Omega_\beta\,{}^t\bar{\Omega}_\beta)\begin{pmatrix} X \\ \bar{X}\end{pmatrix}$; if we put $\sum_{ij}\int_{\mathsf{P}'_{ij}}^{\mathsf{P}_{ij}}(\Phi) = (v^*)$, we have

$$(\chi_{\mathsf{Y}}(\gamma)^e 1) = \exp(2\pi i(v^*\bar{v}^*)\begin{pmatrix} X \\ \bar{X}\end{pmatrix}(eB^{-1})).$$

Now there exists a real vector (m') with $2p$ columns such that $(v^*) = (m')^t\Omega_\beta$; and we have

$$\begin{aligned}(\chi_{\mathsf{Y}}(\gamma)^e 1) &= \exp(2\pi i e(m')\,(BI_\alpha{}^{-1}\,{}^tB)^{-1} \\ &= \exp(2\pi i e(m')\left(\begin{array}{c|c} \boxed{{}^tE_{\mathsf{M}}{}^{-1}} & 0 \\ \hline 0 & \boxed{} \end{array}\right)).\end{aligned}$$

Therefore if we denote by (m) a real vector with $2q$ columns such that $(v) = (m)^t\omega$, we have

$$(\chi_{\mathsf{Y}}(\gamma)^e) = \exp(2\pi i e\epsilon(m)\,{}^tE_{\mathsf{M}}{}^{-1}).$$

Now we may take as z every generic point of A over $K(\mathsf{M})$; such points are everywher dense in A. However since $(\chi_{\mathsf{Y}}(\gamma))$ and $\exp(2\pi i\epsilon(m)\,{}^tE_{\mathsf{M}}{}^{-1})$ both depend continuously on z,[22] and they approach (1), if z approaches z', we have

$$(\chi_{\mathsf{Y}}(\gamma)) = \exp(2\pi i\epsilon(m)\,{}^tE^{-1}).$$

---

[21] Cf. H. Weyl, *Die Idee der Riemannschen Fläche*, Berlin (1913), Kap. II, §§ 16-17. See also J. Igusa, "Zur klassischen Theorie der algebraischen Funktionen," *Journal of the Mathematical Society of Japan*, vol. 1 (1948).

[22] That the first character depend continuously upon z can be proved rigorously by Kodaira's generalization of Weyl's formula: Green's formula and meromorphic functions on compact analytic varieties, to appear in the *Canadian Journal of Mathematics*.

Thereby we write E instead of E_{M}, since it does not depend on the choice of $\mathsf{C}(\mathsf{M})$. In fact this follows from the above formula, if we remember that $(\chi_{\mathsf{Y}}(\gamma))$, ϵ and (m) are independent of $\mathsf{C}(\mathsf{M})$.

10. On the other hand if we replace the period matrix ω by $\hat{\omega} = \omega\epsilon^{-1}E$, $[\omega]$ is a subgroup of index ϵ^{2q-2} in the similar group $[\hat{\omega}]$. Moreover since $\hat{\omega}$ is also a Riemann matrix, there exists an abelian variety $\mathfrak{P}$ in a suitable projective space, which is isomorphic with $S^q/[\hat{\omega}]$ as complex toroids, and a homomorphism λ from A to $\mathfrak{P}$ such that $\lambda\Theta(v) = 0$ if and only if (v) is contained in $[\hat{\omega}]$. If we define a real vector (r) with $2q$ columns by $(v) = (r){}^t\hat{\omega}$, we have $(\chi_{\mathsf{Y}}(\gamma)) = \exp(2\pi i(r))$. Therefore if we denote by $\mathfrak{G}_l(\mathsf{V})$ the group of V-divisors, which are linearly equivalent to 0 in V, then $\mathsf{X}(z_1) \equiv \mathsf{X}(z_2)$ (mod. $\mathfrak{G}_l(\mathsf{V})$) if and only if $\lambda(z_1) = \lambda(z_2)$. This abelian variety $\mathfrak{P}$ is the same as the *Picard variety attached to* V in the sense of the Italian geometers.

It follows from the above results that if a V-divisor Y is homologous to 0 modulo Z, we can find a V-divisor $\mathsf{X}(z) - \mathsf{X}(z')$, which induces the same character on $\mathfrak{B}(\mathsf{V})$ as Y. Then, by the definition of the character, Y is linearly equivalent to $\mathsf{X}(z) - \mathsf{X}(z')$. Since linear equivalence implies continuous equivalence in the sense of § I, 2, and since $\mathsf{X}(z)$ is continuously equivalent to $\mathsf{X}(z')$, by Theorem 1, we see readily that Y is continuously equivalent to 0. As we have remarked in § I, 2, however, continuous equivalence implies the homological equivalence modulo Z; thus we can state

THEOREM 4. *In the case of divisors on the projective model, the continuous equivalence and the homological equivalence* modulo Z *are the same.*[23]

This theorem was first proved by Lefschetz[24] essentially along the same line as above. We shall denote by $\mathfrak{G}_c(\mathsf{V})$ the group of V-divisors, which is

---

[23] If $\mathfrak{G}^r{}_w(\mathsf{V}_\alpha)$ and $\mathfrak{G}^r{}_{w'}(\mathsf{V}_\alpha)$ correspond to the complete variety V_α without multiple point ($r \leqq \dim(\mathsf{V}_\alpha)$) by the "strictest equivalence theories" satisfying (A), (B), (C′), (D), (E), (S) and (A), (D), (E), (S) of (F), Chap. IX, § 7 respectively, we have

$$\mathfrak{G}^{d-1}{}_w(\mathsf{V}) = \mathfrak{G}^{d-1}{}_{w'}(\mathsf{V}) = \mathfrak{G}_c(\mathsf{V}).$$

In fact if $\mathfrak{G}^r{}_h(\mathsf{V}_\alpha)$ corresponds to V_α by the homological equivalence theory over Z, it is an equivalence theory satisfying (A), (B), (C′), (D), (E) and (S); hence

$$\mathfrak{G}^r{}_{w'}(\mathsf{V}_\alpha) \subset \mathfrak{G}^r{}_w(\mathsf{V}_\alpha) \subset \mathfrak{G}^r{}_h(\mathsf{V}_\alpha)$$

for every V_α. Moreover by the same arguments as in § I, 2, we have

$$\mathfrak{G}_c(\mathsf{V}) \subset \mathfrak{G}^{d-1}{}_{w'}(\mathsf{V});$$

and $\mathfrak{G}^{d-1}(\mathsf{V}) = \mathfrak{G}_c(\mathsf{V})$ by Theorem 4.

[24] Cf. loc. cit. 15, Chap. IV.

defined by the continuous equivalence on V, and we shall resume our main results, which include the classical "inversion theorem" of Jacobi, as follows.

THEOREM 5. *We can attach two abelian varieties, the Albanese variety* A *and the Picard variety* $\mathfrak{P}$, *to every non-singular projective model* V. $\mathfrak{P}$ *is obtained by a homomorphism from* A, *which corresponds to the division of the period matrix of the Picard integrals of the first kind on* V *by one of its principal matrices. Moreover,* $\mathfrak{P}$ *is isomorphic with the factor group* $\mathfrak{G}_c(V)/\mathfrak{G}_l(V)$, *and this is dually paired with the Betti group* $\mathfrak{B}(V)$ *by the multiplication rule*

$$\left.\begin{matrix}\mathfrak{G}_c(V)\ni Y\\ \mathfrak{B}(V)\ni\gamma\end{matrix}\right\}\ Y\cdot\gamma=\chi_Y(\gamma).$$

In addition every element of $\mathfrak{G}_c(V)/\mathfrak{G}_l(V)$ has a representative of the form $X(z)-X(z')$ with $\deg. X(z)=\deg. X(z')=q\epsilon$.

On the other hand if we introduce a square *real* matrix J of degree $2q$ by $i\omega=\omega J$, then an isomorphism $\omega\to\Lambda\omega L$ induces an isomorphism $J\to L^{-1}JL$. Conversely ω is determined up to a special isomorphism $\omega\to\Lambda\omega$ by J. This matrix J is explicitly introduced in the theory of complex toroids by Weil [25]; hence it may be called the *Weil matrix attached to* ω. A real matrix J is a Weil matrix if and only if $J^2=-1$ and if there exists a skew-symmetric Q-matrix E such that EJ is symmetric and definite. On the other hand a Q-matrix S satisfies the relation $\omega S^t\omega=0$ if and only if

$$\begin{pmatrix} \boxed{i} & 0\\ 0 & \boxed{-i}\end{pmatrix} \text{ commutes with } \begin{pmatrix}\omega\\ \bar{\omega}\end{pmatrix} S({}^t\bar{\omega}\,{}^t\omega);$$

hence if and only if S satisfies the equation $JS^tJ=S$. In particular we have $JE^tJ=E$. It follows therefore that

$$i\hat{\omega}=i\omega\epsilon^{-1}E=\omega J\epsilon^{-1}E=\omega\epsilon^{-1}E^tJ^{-1}=\hat{\omega}^tJ^{-1}.$$

Thus we get the following result.

COROLLARY. *If* J *is the Weil matrix attached to* ω, *then* ${}^tJ^{-1}$ *is the Weil matrix attached to* $\hat{\omega}$.

Now we shall consider the Picard variety of an abelian variety. First of all, if V^d is an abelian variety, the Albanese variety A^q attached to V is isomorphic with V. In fact there exists a K-matrix ω of type $(d, 2d)$ such that the kernel of the homomorphism h from the universal covering group S^d

[25] A. Weil, "Théorèmes fondamentaux de la théorie des fonctions thêta," *Seminaire Bourbaki* (Mai, 1949).

of V^d onto V coincides with $[\omega]$. Then if we denote by $(v_1, \cdots, v_d)$ the complex coordinates in S^d, their differentials $dv_\alpha = dv_\alpha(h^{-1}(P))$ $(1 \leqq \alpha \leqq d)$ are linearly independent Picard differentials of the first kind on V, which form a base of such differentials since $R^1(V) = 2d$. Moreover, since they have the period matrix ω (along a set of suitable 1-cycles in V), A is isomorphic with $S^d/[\omega]$, hence with V. It follows from this fact and from the above corollary that the Picard variety $\mathfrak{P}$ attached to the given variety V is also attached to the Albanese variety A of V as its Picard variety. Since the operation $J \to {}^tJ^{-1}$ in the same corollary is involutive, we get the following duality of Weil: *The Picard variety of an abelian variety V is not V, but another abelian variety V'; the Picard variety of V' is then V, and there is a kind of duality between them.*

The following figure shows the relations of various varieties, which appear in our theory.

$$\begin{tikzcd}
 & A \arrow[d, shift left, "\text{Picard variety}"] & \text{(its own Albanese variety)} \\
V \arrow[ur, "\text{Albanese variety}"] \arrow[dr, "\text{Picard variety}"'] & & \\
 & \mathfrak{P} \arrow[u, shift left] & \text{(its own Albanese variety)}
\end{tikzcd}$$

11. Until now we have fixed our projective model V; we shall discuss therefore in what manner our theory depends on the choice of it. Let V' be another non-singular projective model, which is equivalent to V by a birational correspondence T. Then there exists a bunch $\mathfrak{B}$ of $(d-1)$ (complex) dimension (at most) on V^d such that T is biregular at every point of $V - \mathfrak{B}$. Moreover there exist a bunch $\mathfrak{F}$ of (complex) dimension at most $(d-2)$ on V such that the projection from T to V is regular at every point of $V - \mathfrak{F}$. Therefore T induces an isomorphism of the Poincaré group of V onto that of V'. This fact was remarked explicitly for the first time by Ehresmann on the last page of his thesis.

On the other hand let $\Phi'(P')$ be any algebraic differental of the first kind on V'. Then the differential $\Phi(P) = \Phi'(T(P))$ on V is also of the first kind. This was proved (explicitly) by Kähler[26]; and the main idea is as follows. If Φ is not of the first kind, it has at least one "pole" Y on V; but since $T(P)$ is regular along Y, $\Phi'(T(P))$ is finite along Y, a contradiction.

---

[26] E. Kähler, "Forme differentiali e funzioni algebriche," *Mem. Accad. Ital. Mat.*, vol. 3 (1932). Cf. also S. Koizumi, "On the differential forms of the first kind on algebraic varieties," *Journal of the Mathematical Society of Japan*, vol. I (1949).

Therefore if ω is the period matrix of the Picard integrals of the first kind on V, it is also a similar matrix for V'. We have thus arrived at the following result.

THEOREM 6. *If* A *and* A' *are the Albanese varieties attached to* V *and* V' *respectively, then they are isomorphic.*

A simple consequence of this theorem and corollary to Theorem 5 is

THEOREM 7. *If* $\mathfrak{P}$ *and* $\mathfrak{P}'$ *are the Picard varieties attached to* V *and* V' *respectively, then they are isomorphic.*

We add two Appendices in which, it is hoped, a fairly complete theory of divisors on a non-singular projective model will be established.

Apendix I. Numerical Equivalence and the Lefschetz Number.

We shall first obtain a duality theorem including the torsion cycles. By a duality of Poincaré, the 1-dimensional torsion group of V^d is isomorphic with the $(2d-2)$-dimensional torsion group of V. This duality is based on the fact that we can select the "dual bases" $\delta_1, \cdots, \delta_t$ and $\mathsf{Y}_1, \cdots, \mathsf{Y}_t$ of the torsion groups of 1 and $(2d-2)$-dimensions respectively such that

$$L(\delta_i, \mathsf{Y}_j; \mathsf{V}) = \delta_{ij}/m_j \qquad (1 \leqq i, j \leqq t),$$

where m_j means the common order of δ_j and Y_j for $1 \leqq j \leqq t$. Moreover by the Lefschetz-Hodge theorem in (H), § 51. 2, we may assume that Y_i are all (algebraic) V-divisors. Since Y_i is homologous to 0 modulo Q, there exists a multiplicative function F_i on V such that $(F_i) = \mathsf{Y}_i$ for $1 \leqq i \leqq t$; and we have

$$\chi_{\mathsf{Y}_i}(\delta_j) = \exp(2\pi i \delta_{ij}/m_j) \qquad (1 \leqq i, j \leqq t).$$

Therefore if we denote by $\mathfrak{G}_n(\mathsf{V})$ the group of V-divisors, which are homologous to 0 modulo Q, we have the following *duality theorem*.

The factor group $\mathfrak{G}_n(\mathsf{V})/\mathfrak{G}_l(\mathsf{V})$ is dually paired with the homology group $\mathfrak{H}^1(\mathsf{V}, Z)$ by the multiplication rule

$$\left.\begin{matrix} \mathfrak{G}_n(\mathsf{V}) \ni \mathsf{Y} \\ \mathfrak{H}^1(\mathsf{V}, Z) \ni \Gamma \end{matrix}\right\} \mathsf{Y} \cdot \Gamma = \chi_{\mathsf{Y}}(\Gamma).$$

It follows that the group $\mathfrak{G}_n(\mathsf{V})/\mathfrak{G}_l(\mathsf{V})$ does not depend on the choice of the projective model V.

Now we shall show that the group $\mathfrak{G}_n(V)$ coincides with the set of V-divisors, which are *numerically equivalent to* 0. A V-divisor Y is defined to be numerically equivalent to 0, if it satisfies $I(Y, \Gamma; V) = 0$ for every algebraic 1-cycle Γ in V. Since the Kronecker index is always defined, the numerical equivalence defines a group of V-divisors. We note also that the homological equivalence modulo Q is broader than the continuous equivalence and is stricter than the numerical equivalence.

Let $\mathfrak{G}(V)$ be the group of all V-divisors, then the factor group $\mathfrak{G}(V)/\mathfrak{G}_n(V)$ has a finite number $\rho(V)$ of independent generators $Z_1, \cdots, Z_\rho$ such that $\rho(V) \leqq R^{2d-2}(V) = R^2(V)$. Moreover by Proposition 3, if W^2 is a generic 2-section of V over a common field of definition of $Z_1, \cdots, Z_\rho$, then $\Gamma_i = Z_i \cdot W^2$ $(1 \leqq i \leqq \rho)$ are independent algebraic 1-cycles in V modulo Q. It can be readily seen by Lefschetz's theorem and by Hodge's extension of Poincaré's theorem in (H), § 50 that we can find a set of algebraic 1-cycles Γ'_j on the surface W, which are homologous to 0 in V modulo Q, such that Γ_i and Γ'_j form a base of algebraic 1-cycles on W modulo Q and such that $I(\Gamma_i, \Gamma'_j; W) = 0$. Then it follows from Severi's theorem that the intersection matrix

$$(I(Z_i, \Gamma_j; V)) = (I(\Gamma_i, \Gamma_j; W))$$

is non-singular. On the other hand any V-divisor Y can be written uniquely in the form

$$Y \sim \sum_{i=1}^{\rho} a_i Z_i \qquad \text{(in V mod. } Q)$$

with Z-coefficients a_i $(1 \leqq i \leqq \rho)$. If Y is numerically equivalent to 0, so is $\sum_{i=1}^{\rho} a_i Z_i$, hence by what we have just proved, this must be 0; which proves the assertion.

The integer $\rho(V)$ is called the *Picard number of* V^d; we shall call the difference $\tau(V) = R^{2d-2}(V) - \rho(V)$ the *Lefschetz number* of V, and we shall prove its absolute invariance. Let $Z_1, Z_2, \cdots, Z_\tau$ be a base of the "transcendental" $(2d-2)$-cycles in V, then $\Gamma_i = Z_i \cdot W^2$ $(1 \leqq i \leqq \tau)$ form a base of the transcendental 2-cycles in V. Therefore by a similar argument as in (H), § 51. 2, we have, for every algebraic differential of degree 2 of the first kind Φ in V,

$$\int_{\sum_{i=1}^{\tau} a_i \Gamma_i} \Phi = 0,$$

with a_i in Z, if and only if $a_i = 0$ $(1 \leqq i \leqq \tau)$. On the other hand let T be a birational correspondence between V and V'. Then we may speak of the image Γ'_i of Γ_i by T for $1 \leqq i \leqq \tau$; and we have

$$\int_{\sum_{i=1}^{\tau} a_i \Gamma'_i} \Phi' = 0$$

for every algebraic differential of degree 2 of the first kind Φ' in V', if and only if $a_i = 0$ $(1 \leqq i \leqq \tau)$. Therefore $\Gamma'_1, \cdots, \Gamma'_\tau$ are independent transcendental 2-cycles in V'. Accordingly we must have $\tau(\mathsf{V}) \leqq \tau(\mathsf{V}')$, and similarly $\tau(\mathsf{V}') \leqq \tau(\mathsf{V})$, which completes the proof.

Appendix II. Algebraic Correspondence and Poincaré Family.

Let $\mathsf{V}_1^{d_1}$ and $\mathsf{V}_2^{d_2}$ be two non-singular projective models with irregularities q_1 and q_2 respectively, and consider the module of $(\mathsf{V}_1 \times \mathsf{V}_2)$-divisors over Z; its member may be called a *correspondence between* V_1 *and* V_2. It can be readily seen that a correspondence X satisfies $\mathsf{X}(z_1) \equiv \mathsf{X}(z_2)$ (mod. $\mathfrak{G}_l(\mathsf{V}_2)$), for every pair of points z_1 and z_2 on V_1, if and only if X is of the form

$$\mathsf{X} \equiv \mathsf{V}_1 \times \mathsf{Y}_2 + \mathsf{Y}_1 \times \mathsf{V}_2 \quad (\text{mod. } \mathfrak{G}_l(\mathsf{V}_1 \times \mathsf{V}_2)),$$

with some V_α-divisors Y_α $(\alpha = 1, 2)$. We shall call such an X a *correspondence with valence* 0. Since the correspondences with valence 0 form a submodule of the module of all correspondences, we can consider the factor module $\mathfrak{B}(\mathsf{V}_1, \mathsf{V}_2)$. On the other hand since V_α are orientable manifolds, $\mathfrak{H}^{2d_\alpha - 1}(\mathsf{V}_\alpha, Z)$ are free abelian groups with $2q_\alpha$ generators $\Gamma_{\alpha i}$ $(1 \leqq i \leqq 2q_\alpha)$; and if we put $\Gamma_{\alpha i} \cdot \mathsf{W}_\alpha^1 = \gamma_{\alpha i}$ $(1 \leqq i \leqq 2q_\alpha)$, where W_α^1 are generic 1-sections of V_α, then the $\gamma_{\alpha i}$ $(1 \leqq i \leqq 2q_\alpha)$ form a base of $\mathfrak{H}^1(\mathsf{V}_\alpha, Q)$ for $\alpha = 1, 2$. Moreover every $2(d_1 + d_2 - 1)$-cycle X of $\mathsf{V}_1 \times \mathsf{V}_2$ over Z can be written uniquely in the form

$$X \sim \mathsf{V}_1 \times \Gamma_2 + \Gamma_1 \times \mathsf{V}_2 + \sum_{ij} s_{ij}(\Gamma_{1i} \times \Gamma_{2j}) \ (\text{mod. } Z),$$

where Γ_α are Z-cycles of $2(d_\alpha - 1)$ dimension in V_α $(\alpha = 1, 2)$ and where $S = (s_{ij})$ is a Z-matrix of type $(2q_1, 2q_2)$. Let ω_α be the period matrices of the Picard integrals of the first kind on V_α along the 1-cycles $\gamma_{\alpha i}$ $(1 \leqq i \leqq 2q_\alpha)$ and let A_α be the abelian varieties attached to these Riemann matrices for $\alpha = 1, 2$. It can be readily calculated without difficulties by the Lefschetz-Hodge theorem that X is algebraic if and only if Γ_α are algebraic and S satisfies the relation $\omega_1 S^t \omega_2 = 0$. Therefore if we denote by $\lambda(\mathsf{V}_1, \mathsf{V}_2)$ the

814712

rank of the Z-module of such matrices S, we have the following formula, which is well known if V_1 and V_2 are curves:

$$\rho(V_1 \times V_2) = \rho(V_1) + \rho(V_2) + \lambda(V_1, V_2).$$

Now the module $\mathfrak{H}(A_1, A_2)$ of all homomorphisms of A_1 into A_2 can be represented faithfully as the module of "complex multiplications" of ω_1 to ω_2; and if we extend this module over Q, it is isomorphics with the module of all Q-matrices satisfying $\omega_1 S^t \omega_2 = 0$. Moreover if a correspondence X is homologous to 0 modulo Z, we can conclude from Theorem 5 that X is of valence 0. Therefore the Z-modules $\mathfrak{B}(V_1, V)$ and $\mathfrak{H}(A_1, A_2)$ have the same rank; since they are both free, however, we get the isomorphism

$$\mathfrak{B}(V_1, V_2) \cong \mathfrak{H}(A_1, A_2).$$

As a free abelian group, $\mathfrak{H}(A_1, A_2)$ depends only on the categories of A_1 and A_2.

If we apply the above result to the case where $V_1 = A$ is the Albanese variety attached to a given variety $V_2 = V$. we see that $\mathfrak{B}(A, V)$ is generated by the Poincaré family in the "general case" $\mathfrak{H}(A, A) = Z$; and then, up to a scalar factor, the Z-matrix E in our theory is an absolute invariant of V. In this connection it is also to be remarked that there exists a Riemann matrix with any preassigned principal matrix; thus the Albanese variety and the Picard variety attached to the same variety are not isomorphic in general.

KYOTO UNIVERSITY, JAPAN.

CONTOUR EQUIVALENT PSEUDOHARMONIC FUNCTIONS AND PSEUDOCONJUGATES.*

By James A. Jenkins and Marston Morse.

1. Introduction. The present paper deals with the structure of the level curves of pseudoharmonic functions defined in the closure of Jordan domains D and satisfying certain boundary conditions. We are in particular interested in formulating necessary and sufficient conditions that two such functions U and V be *contour equivalent* (written C. E.) that is, that there exist a sense-preserving homeomorphism (written S. P. homeomorphism) ϕ of the domain of U onto that of V under which the connected level arcs of U are mapped onto the connected level arcs of V. By defining and constructing a pseudoconjugate v of an admissible u an interior transformation $u + iv$ of D is obtained. It is proved that for a given admissible U a S. P. homeomorphism ϕ of $\bar{D}$ onto itself exists such that the composite function $U\phi$ is harmonic. The theorems on contour equivalence are new even when U and V are both harmonic.

The set of level curves of a pseudoharmonic function U which emanate from the multiple points of U is termed the *net* of U. We show that each admissible U is C. E. to a model U_0 with a net composed of hyperbolic lines (regarding the disc $|z| \leq 1$ as a hyperbolic plane).

Reference will frequently be made to the book: Marston Morse, "Topological Methods in the Theory of Functions of a Complex Variable," Princeton University Press, which we denote by M.

When a pseudoharmonic function U is defined over a general simply connected *open* domain a pseudoconjugate of U still exists although radically different methods of proof are required. It is even possible to start with families of curves which have the topological properties of level arcs but which are not given as level arcs of a function U. Contact is thus made with some of the results of W. Kaplan [1] although the methods used are quite distinct. Results in this *open* case will be presented later.

In the bibliography, reference is made to two additional papers, of interest in connection with this paper.

* Received February 12, 1951.

[1] W. Kaplan, "Topology of level curves of harmonic functions," *Transactions of the American Mathematical Society*, vol. 63 (1948), pp. 514-522.

Part I. Contour Equivalence.

2. Fundamental definitions. A function U with real values $U(z)$, $z = x + iy$, is said to be pseudoharmonic at the point z_0 of the z-plane if $U(z)$ is defined in a neighborhood N of z_0 and if there exists a S. P. homeomorphism ϕ from a neighborhood N_1 of z_0 to N, leaving z_0 fixed, such that the composite function $U\phi$ is harmonic and non-constant on N_1. A critical point of $U\phi$ is termed a *critical* point of U. A function U is said to be pseudoharmonic on a domain D (open) if it is pseudoharmonic at each point of D. A subset of D on which $U(z)$ is contant will be called a *U-set*. A *U-arc* is thus well-defined. We understand that a U-set but not a U-arc may reduce to a point. We shall introduce canonical neighborhoods N of a point $z_0 \varepsilon N$ and canonical representations of U over N as follows.

Case I. Let U be pseudoharmonic in a neighborhood of z_0. There then exists a neighborhood N of z_0 free from critical points of U except at most z_0, with $\bar{N}$ the homeomorph of a plane circular disc such that z_0 corresponds to the center of the disc and the locus on which $U(z) = U(z_0)$ in N corresponds to a set of $2n$ rays $(n > 0)$ leading from this center and making successive sectors of central angle π/n. As a variable point z crosses any one of these U-arcs (except at z_0) the difference $U(z) - U(z_0)$ changes sign. (See M., Th. 2.1.) If $n = 1$, a single U-arc passes through z_0 and z_0 is termed *ordinary*. If $n > 1$ z_0 is termed a *multiple* point of U of *index* $2n - 2$. A neighborhood of z_0 such as N will be termed *canonical*.

Let D be a Jordan domain bounded[2] by a Jordan curve βD.

Conditions[3] $[D]$. *A real valued function U defined over $\bar{D}$ will be said to be in $[D]$ if U is pseudoharmonic on D, continuous on $\bar{D}$, and if $U \mid \beta D$ assumes its relative extreme values in at most a finite set of points in βD.*

We recall a number of properties of a $U \varepsilon [D]$. Every point $z_0 \varepsilon \beta D$ has a canonical neighborhood coming under Case II or Case III.

Case II. $z_0 \varepsilon \beta D$; *z_0 not a point of relative extremum of U.* Here z_0 has a canonical vicinity N relative to $\bar{D}$, free from critical points (cf. M) of $\pm U$ in $\bar{D}$, with $\bar{N}$ the homeomorph of a semi-disc H, such that z_0 corresponds to the center O of H, $\bar{N} \cap \beta D$ corresponds to the diameter of H, while the U-set in N at the level $c = U(z_0)$ is represented by n rays $(n > 0)$ emanating

[2] As a general notation, βD shall mean the boundary of D.

[3] Conditions $[D]$ include conditions called Boundary Conditions A in Morse, loc. cit.

from O, and dividing N into $m = n + 1$ sectors (open) on which $U - c$ alternates in sign. If $m > 2$, z_0 is termed a boundary *multiple* point of U of *index* $m - 1$.

Case III. $z_0 \varepsilon \beta D$; z_0 *a point of relative extremum of* U. In this case there exists a canonical neighborhood N of z_0 relative to $\bar{D}$ free from critical points of $\pm U$ with $\bar{N}$ the homeomorph of a semi-disc such that z_0 corresponds to the center of the semi-disc, $\bar{N} \cap \beta D$ corresponds to the diameter of the semi-disc, the boundary of N in D is at a level $c_1 \neq c$ such that for $z \varepsilon N$ and $z \neq z_0$, $U(z)$ is between c_1 and c. (M., Th. 7.2.) The point z_0 is termed *ordinary* in this case.

U-continuation. As shown in M., § 7, a $U \varepsilon [D]$ has only a finite number of multiple points in $\bar{D}$. By a U-continuation of a simple U-arc g will be meant a simple U-arc containing g which continues through a multiple point z_0 of U in D with the U-arc which issues from z_0 opposite to g in the canonical representation of U neighboring z_0.

Contour Equivalence. *Let* Δ *be a second Jordan domain. Two functions* U *and* V *continuous over* $\bar{D}$ *and* $\bar{\Delta}$ *respectively, are said to be* C. E. *under* ϕ, *if there exists a* S. P. *homeomorphism* ϕ *of* $\bar{D}$ *onto* $\bar{\Delta}$ *such that each maximal connected U-set (V-set) corresponds to a like V-set (U-set)*; ϕ *is said to define a strict contour equivalence if* $U = V\phi$.

Examples show that not every contour equivalence of pseudoharmonic functions is strict. However, we shall prove in § 7 that a contour equivalence of a $U \varepsilon [D]$ with a $V \varepsilon [D]$ implies a strict contour equivalence, if one admits a preliminary transformation $U' = -U$ (if necessary) and a preliminary continuous deformation of U through functions in $[D]$ with fixed level arcs. One sees that contour equivalence, as well as strict contour equivalence, is symmetric, reflexive, and transitive.

Type of multiple point. In § 4 we shall enlarge the class $[D]$ to the class $[D]'$. Understanding the term multiple point P and canonical neighborhood $N(P)$ in the enlarged sense of § 4 as well as in the sense of § 2, let P and Q be multiple points of U with canonical neighborhoods $N(P)$ and $N(Q)$. We term P and Q multiple points of the *same* type if for some choice of $N(P)$ and $N(Q)$, $\bar{N}(P)$ admits a homeomorphism onto $\bar{N}(Q)$ which maps U-arcs onto V-arcs.

It is readily seen that multiple points of the same type have equal indices and both come under the same one of the Cases I, II (or V, VII in the enlarged sense).

Interior transformations f. A function f with complex values $f(z)$ in the $w = u + iv$-plane will be said to be *interior at* z_0 if f is defined in a neighborhood N of z_0, and if there exists a S. P. homeomorphism ϕ of a neighborhood N_1 of z_0 onto N leaving z_0 fixed and such that the composite function $f\phi$ is meromorphic and non-constant on N_1. We term f *interior* over D if f is interior at each point $z_0 \varepsilon D$. If f is interior over D the real and imaginary parts of f are pseudoharmonic over D apart from poles of f.

The following two lemmas will be useful in § 11, § 12.

LEMMA 2. 1. *Let* z_0 *be a point in* D. *If* f *is interior in* $D - z_0$ *and continuous in* D, *then* f *is interior at* z_0.

Set $f(z_0) = w_0$. For $e > 0$ sufficiently small, the set $E_e: |z - z_0| \leqq e$ is in D and for some such e the image curve $g = f(\beta E_e)$ does not intersect w_0. Let e be so chosen. Let m be the order of g with respect to w_0. It is seen that each point $w \neq w_0$ in a sufficiently small neighborhood of w_0 is covered m times by $f \mid (E_e - z_0)$, so that f is open and light. The lemma follows.

LEMMA 2. 2. *If* f *is interior in* D, *continuous in* $\bar{D}$ *and if* $f(\beta D) = g$ *is locally simple, then* f *admits an interior extension over some neighborhood of any given point* z_0 *of* βD.

As seen in M., p. 85 the points $z_0 \varepsilon \beta D$ neighboring which f fails to be topological are isolated in βD. Lemma 2. 2 is immediate except at a point z_0 neighboring which f fails to be topological. Suppose then that z_0 is such a point.

As seen in M., § 23 one can suppose without loss of generality that $z_0 = 0$ and that there is an open arc h of βD of the form $[y = 0, -e < x < e]$ with an image $f(h)$ covering an arc $[v = 0, a < u < b]$ in a 1-1 manner. We continue f over h by requiring that $f(\bar{z}) = \bar{f}(z)$ for points $z \varepsilon \bar{D}$ near $z = 0$. It is seen that the extended f satisfies the conditions of Lemma 2. 1 in a sufficiently small neighborhood of $z_0 = 0$ and so is interior at z_0.

3. Conditions necessary for contour equivalence. Contour equivalence of $U \varepsilon [D]$ with $V \varepsilon [\Delta]$ under ϕ implies a 1-1 mapping T of the set of multiple points of U onto the set of multiple points of V, sending a multiple point of U into a multiple point of V of the same type, and boundary multiple points into boundary multiple points in the same cyclic order. If a finite 1-1 correspondence T of this type is given without giving ϕ but satisfying certain other finite conditions to be enumerated, there then exists (§ 6) an

extension ϕ of T over $\bar{D}$ such that ϕ defines a contour equivalence of U with V. New terms are needed.

The star $S(P, U)$. Let P be a multiple point of U in D or βD. Let each U-arc g with end point at P be "continued" (§ 2) on D (open) from P until a point of βD is reached. So continued and sensed g will be termed a *ray* of the *star* $S(P, U)$ of P. The rays of $S(P, U)$ will be simple and non-intersecting except at P. They will divide D into open connected sectors

$$S_1(P, U), \cdots, S_m(P, U) \qquad (m > 2) \tag{3.1}$$

where m is even when $P \varepsilon D$. We suppose that these sectors are indexed in the order in which their boundary arcs on βD follow each other in counter-clockwise sense, starting with the first such arc of βD following P in case $P \varepsilon \beta D$. For $P \varepsilon \beta D$, $S_1(P, U)$ will then be uniquely determined. For $P \varepsilon D$, $S_1(P, U)$ can be taken as any one of the sectors of $S(P, U)$. The boundary of $S_i(P, U)$ will include an arc of βD and one or two rays of $S(P, U)$. We term P the *center* of the star $S(P, U)$. Two stars will be said to be of the same *type* if their centers are of the same type.

Frames $M(U)$ *and mappings* T. We shall obtain homeomorphisms ϕ defining contour equivalence of pseudoharmonic functions U and V as extensions over $\bar{D}$ of mappings T of special subsets (§ 4) $M(U) \subset \bar{D}$. To be admissible a *frame* $M(U)$ shall include all the multiple points of U. A mapping T of $M(U) \subset \bar{D}$ onto $M(V) \subset \bar{\Delta}$ to be admissible must be a homeomorphism which preserves the type of each multiple point and which maps $M(U) \mid \beta D$ onto $M(V) \mid \beta\Delta$ with preservation of sense in case $M(U) \mid \beta D$ contains at least three points. With this understood we define similarity of $M(U)$ with $M(V)$ under T relative to two stars $S(P, U)$ and $S(Q, V)$ for which $Q = TP$.

RELATIVE SIMILARITY UNDER T. *Let T be an admissible mapping of $M(U)$ onto $M(V)$ with a multiple point P of U going into a multiple point Q of V of the same type. We say that $[M(U), S(P, U)]$ is similar to $[M(V), S(Q, V)]$ under T if for some admissible indexing of $S(P, U)$ and $S(Q, V)$ and for each point $z \varepsilon M(U)$ and image $w \varepsilon M(V)$ under T the incidence relations*

$$z \varepsilon \bar{S}_i(P, U), \qquad w \varepsilon \bar{S}_i(Q, V) \qquad (i = 1, \cdots, n) \tag{3.2}$$

both hold or both fail to hold for each i.

SIMILARITY UNDER T. *Admissible sets $M(U) \subset \bar{D}$ and $M(V) \subset \bar{\Delta}$, homeomorphic under an admissible mapping T, are termed similar under T if for any two multiple points P and Q which are images under T, $[M(U), S(P,U)]$ is similar to $[M(V), S(Q,V)]$ under T.*

SIMILARITY. *Admissible sets $M(U) \subset \bar{D}$ and $M(V) \subset \bar{\Delta}$ are termed similar (written $M(U) \sim M(V)$) if $M(U)$ and $M(V)$ are similar under some admissible T.*

It is clear that the relation of similarity between admissible frames $M(U)$ and $M(V)$ is reflexive, symmetric, and transitive. It is thereby not excluded that $U = V$ and that $M(U) \sim M(U)$ under some T other than the identity.

The basic conditions for contour equivalence are given in Th. 3. 1. That these conditions are sufficient will be proved in § 6.

THEOREM 3. 1. *In order that $U \varepsilon [D]$ and $V \varepsilon [\Delta]$ be C. E. under ϕ it is necessary and sufficient that the set $\mu(U)$ of multiple points of U and the set $\mu(V)$ of multiple points of V be similar under some admissible mapping T.*

The condition is necessary, since the existence of the mapping ϕ implies the existence of the mapping $T = \phi \mid \mu(U)$ of $\mu(U)$ onto $\mu(V)$. Moreover, T is admissible in that T, like the mapping ϕ, preserves the type of the multiple point, carries boundary multiple points into boundary multiple points and preserves cyclic order of these boundary points.

In satisfying the condition of similarity of $\mu(U)$ with $\mu(V)$ an indexing (3. 1) of the sectors of $S(P,U)$ must be made and coordinated with an indexing of the sectors of $S(Q,V)$ where $T(P) = Q$ so that the incidence conditions (3. 2) both hold or fail to hold. When P and Q are boundary points only one indexing is possible. If P and Q are in D and happen to be the only multiple points of U and V respectively, Th. 3. 1 is vacuous except for the condition that P and Q have the same multiplicity; in this case the sectors $S_1(P,U)$ and $S_1(Q,V)$ can be chosen arbitrarily from among the sectors of $S(P,U)$ and $S(Q,V)$ respectively.

If U possesses a multiple point P' different from P, and if V is C. E. with U, then V possesses the multiple point $T(P') = Q' \neq Q$. Given an indexing (3. 1) of the sectors of $S(P,U)$, an indexing of the sectors of $S(Q,V)$ is then uniquely determined by the condition that the sector (or two adjacent sectors) of $S(P,U)$ with whose closure (closures) P' is incident, bear the same index (indices) as the sector (or two adjacent sectors) of $S(Q,V)$ with whose closure (closures) Q' is incident.

4. Conditions $[D]'$. To establish the sufficient conditions for contour equivalence of U and V as given in Th. 3.1 it is convenient to establish a similar theorem in which the preceding class $[D]$ is enlarged. In § 12 we shall show that a $U \varepsilon [D]$ admits a pseudoharmonic continuation over an open domain $D' \supset \bar{D}$. If a U were constant on some arc of βD but otherwise satisfied the conditions $[D]$ this pseudoharmonic continuation over βD might prove impossible, as examples [4] would show. This fact motivates the following definition.

CONDITIONS $[D']$. *A function U defined and continuous at each point of $\bar{D}$ will be said to be in $[D]'$, if U is pseudoharmonic over D, if there are at most a finite set of maximal connected U-arcs in βD, if $U \mid \beta D$ assumes its relative extrema in at most a finite set of such U-arcs or points in βD, and if U admits a pseudoharmonic continuation over βD.*

Suppose $U \varepsilon [D]'$. If z_0 is in a maximal connected U-arc g in βD, a canonical neighborhood N of z_0 and local representation of U over N will not come under Cases I, II, III of § 2. The local representations described below are a ready consequence of the assumption that U admits a pseudoharmonic continuation over βD. There are four new cases.

IV. z_0 interior to g, and a point of relative extremum [5] of U.

V. z_0 interior to g and *not* a point of relative extremum of U.

VI. z_0 an end point of g, and a point of relative extremum of U.

VII. z_0 an end point of g, and *not* a point of relative extremum of U.

In a canonical neighborhood N (relative to $\bar{D}$) of a point z_0 a *sector* of N shall be understood as any maximal connected subset of N on which $U(z) \neq U(z_0)$.

Case IV. A canonical N exists with $\bar{N}$ the homeomorph of a semi-disc H such that z_0 corresponds to the center O of H, $\beta N \cap \beta D$ corresponds to the diameter d of H, and maximal U-arcs in N correspond to chords in H parallel to d. Cf. Case III, § 2.

Case V. An N exists essentially as in Case II except that *all* arcs

[4] One could construct an example in which a U-arc in D had every point of a U-arc in βD as limit point.

[5] That is, in some neighborhood of z_0 relative to $\bar{D}$ $U(z) \geqq U(z_0)$, or else $U(z) \leqq U(z_0)$.

bounding the sectors of N and emanating from z_0 are U-arcs (i. e., including the two arcs of βD emanating for z_0).

Case VI. An N exists homeomorphic with a semi-disc H, such that z_0 corresponds to the center O of H, and $\beta N \cap \beta D$ corresponds to the diameter d of H. Of the two rays ρ_1, ρ_2 of d separated by O, U has its extreme value c at each point of one, say ρ_1, and is strictly monotone on the other. An arbitrary maximal U-arc in N not at the level c corresponds to a quarter circle which starts at a point of ρ_2, has O as its center, and is continued by a straight arc in H parallel to d until βH is reached.

Case VII. An N exists similar to the N under Type V, except that *one* of the sectors of N (the initial or final sector) has one boundary arc in βD, not a U-arc.

Proofs of these statements can be made with complete rigor, using the representations in Cases I, II, and III of the pseudoharmonic extension of U over βD.

Theorem 4. 1 depends for its meaning on the definition of an admissible mapping T of § 3, and this in turn depends upon the meaning attached to two multiple points being of the *same type*. The necessary definitions will now be given.

A canonical neighborhood N of a point $z_0 \varepsilon \beta D$ contains just one sector in Cases III, IV, and VI, and in each of these cases is termed *ordinary*. Let m be the number of sectors in a canonical neighborhood of z_0. In Cases V or VII, $m > 1$ and z_0 is termed a boundary *multiple* point of index $m - 1$. It is seen that a point z_0 of intersection of a U-arc in βD with any U-arc not in βD (except for z_0) always comes under Case V or VII, and so is always a multiple point of positive index.

Given $U \varepsilon [D]'$ and $V \varepsilon [\Delta]'$ let P and Q be multiple points of U and V respectively. For P and Q to be multiple points of the *same type* (as defined in § 3) it is necessary and sufficient that P and Q come under the same one of the Cases I, II, V, or VII and have equal positive indices.

As in § 3 an admissible *frame* $M(U)$ must contain each multiple point of U. Stars $S(P, U)$ are formally defined as in § 3 admitting the new types of multiple points P. As previously two stars $S(P, U)$ and $S(Q, V)$ are termed of the *same type* if P and Q are of the same type. Recall that an admissible mapping T of a frame $M(U)$ onto a frame $M(V)$ is a homeomorphism mapping the set of multiple points of U onto the set of multiple points of V, preserving the type of a multiple point and mapping $M(U) \mid \beta D$ onto $M(V) \mid \beta\Delta$ with preservation of sense on the respective boundaries.

The conditions necessary and sufficient for contour equivalence of U and V include those of Th. 3.1 and are given as follows.

THEOREM 4.1. *Given $U \varepsilon [D]'$ and $V \varepsilon [\Delta]'$ let $M(U)[M(V)]$ be the union of the multiple points of $U[V]$ with U-arcs [V-arcs] in βD, $[\beta\Delta]$. In order that U and V be* C. E. *under a* S. P. *homeomorphism ϕ it is necessary and sufficient that $M(U)$ be similar to $M(V)$ under an admissible mapping T of $M(U)$ onto $M(V)$. Given T, ϕ can be taken as an extension of T over $\bar{D}$. Given ϕ, T can be taken as $\phi \mid M(U)$.*

That the conditions of the theorem are necessary is an immediate consequence of the definitions. That the conditions are sufficient will be established in the next sections. In the next section certain intuitive notions needed in the proofs are made precise.

5. Local right sets, and sensing of U-arcs. Let g be a sensed Jordan arc and $z = a$ an inner point of g. We shall make the intuitive notion of the right of g near a more precise. To that end let N be a Jordan region containing a whose intersection with g is a subarc g' of g forming a cross cut of N. There then exists a S. P. homeomorphism H of N onto the disc $(|z| < 1)$ such that g' is mapped onto the positively sensed segment of the real axis in the disc. The inverse image in N of the set $(|z| < 1, y < 0)$ [alternatively $|z| < 1, y > 0$,] will be called a *local right* [*left*] set of the element (a, g). It is readily shown that for a given element (a, g) local right sets and local left sets of sufficiently small diameter have an empty intersection, while the intersection of any two right (left) sets includes a right (left) set. S. P. homeomorphisms which (by convention) preserve the sense of an arc g, carry right (left) sets of an element (a, g) into right (left) sets of the image element.

Let $z = a$ be an ordinary point of a pseudoharmonic function U. There exists a S. P. homeomorphism ϕ of a neighborhood of $z = a$ into the w-plane $(w = u + iv)$ carrying $z = a$ into $w = 0$ and U-arcs into curves on which v is constant. If g is a properly sensed U-arc through $z = a$, a right set (left set) of (a, g) of sufficiently small diameter thus consists of the points z in a neighborhood of a such that $U(z) < U(a)$ $(U(z) > U(a))$. Such a sensing of g will be termed *U-positive*, the opposite sensing *U-negative*. Such a U-positive sensing of a U-arc g near an ordinary point $z = a$ in g is independent of the mappings and neighborhoods used to define their sensing. If U is pseudoharmonic in D it is possible to assign a positive sense to each U-arc composed of ordinary points of U such that this assignment agrees

with each local assignment of a U-positive sense to a U-arc. At multiple points of U, this sensing is ambiguous.

A particular application of this is as follows. Let a $U \varepsilon [D]'$ assume its absolute minimum on an arc p of βD. Then a U-positive sensing of p will be counter-clockwise on βD. If however U assumes its absolute maximum on an arc q of βD, a U-positive sensing of q will be clockwise.

6. Sufficient conditions for contour equivalence. We shall prove that the conditions of Th. 4.1 are sufficient. Let $\nu(U)$ be the sum of the indices of the multiple points of U, and let $\nu(V)$ be similarly defined. We shall make an induction depending upon the value of $\nu(U) = \nu(V)$ recalling that $\nu(U) = \nu(V)$ when the conditions of Th. 4.1 are satisfied. It is necessary to set $\nu(U) = 0$ if U has no multiple points, similarly for $\nu(V)$. By convention let each ordinary point have an index 0.

LEMMA 6.1. *The truth of* Th. 4.1 *for* $\nu(U) = \nu(V) < n$, *and* $n > 0$ *implies its truth for* $\nu(U) = \nu(V) = n$.

Arcs p and q. It is given that the frames $M(U)$ and $M(V)$ appearing in Th. 4.1 are similar under a homeomorphism T. Let P_0 and Q_0 be fixed multiple points of U and V respectively, corresponding under T. The conditions for star similarity under T of $[M(U), S(P_0, U)]$ and $[M(V), S(Q_0, V)]$ as defined in § 3 are satisfied (by hypothesis) after a suitable indexing of the sectors of $S(P_0, U)$ and a corresponding indexing of the sectors of $S(Q_0, V)$. By virtue of this indexing each ray of $S(P_0, U)$ corresponds to a definite ray of $S(Q_0, V)$. Let p be an arbitrary ray of $S(P_0, U)$ in case $P_0 \varepsilon \beta D$, and in case $P_0 \varepsilon D$ let p be the U-continuation in both senses to βD of an arbitrary ray of $S(P, U)$ with p sensed as the ray. Let q be the corresponding ray of $S(Q_0, V)$, or its continuation in case $Q_0 \varepsilon \Delta$.

By virtue of the assumed similarity of $M(U)$ and $M(V)$ under T, and the choice of p and q, the multiple points of U in p correspond in a 1-1 manner under T to the multiple points of V in q with the order in which the multiple points on p appear on p as a sensed arc the same as the order in which their T-images appear on q as a sensed arc. One sees this on considering incidences with the stars $S(P, U)$ and $S(Q, V)$ where P is an arbitrary multiple point on p and $T(P) = Q \varepsilon q$.

A simplifying modification. Without any loss of generality we can suppose that $\bar{D}$ is the disc $(|z| \leqq 1)$; for under a S. P. homeomorphism ϕ of this disc onto $\bar{D}$, U is C. E. with $U\phi$ (defined over the disc).

The proof can accordingly be simplified by supposing that $\bar{D}$ and $\bar{\Delta}$ are each the disc $(|z| \leqq 1)$, that p and q are each the diameter d of this disc leading from $z = -1$ to $z = +1$, and that the multiple points z^* of U in d are identical with those of V in d, with $T(z^*) = z^*$. The point $z = 1$, if in $M(U)$, is a multiple point of U of Case II, V or VII; likewise the point $z = -1$. Granting this a priori simplification, we assume as previously that $M(U) \sim M(V)$ (under T).

Let D_1 be the open upper semi-disc of D: $(|z| < 1)$. Set $u = U \mid \bar{D}_1$, $v = V \mid \bar{D}_1$. Let $M(u)\,[M(v)]$ be the union of the multiple points of u, $[v]$ with the u-arcs [v-arcs] in βD_1.

The mapping T_1. With u and $v \varepsilon [D_1]'$ an admissible mapping T_1 (see § 3) of $M(u) \subset \bar{D}_1$ onto $M(v) \subset \bar{D}_1$ is obtained on setting $T_1(z) = T(z)$ for $z \varepsilon M(U) \cap \bar{D}_1$ and $T_1(z) = z$ for $z \varepsilon d$. These two conditions on $T_1(z)$ both apply to a point z of $M(U) \cap d$, and are consistent. This is true if $z \varepsilon d$ is a multiple point of U, since $T(z) = z$, as arranged in the preceding paragraphs. It is true if $z = \pm 1$ is in $M(U)$, since this can happen only if $z = \pm 1$ is a multiple point of U [Cf. Cases V and VII of § 4.] The mapping T_1 of the frame $M(u)$ onto the frame $M(v)$ is admissible in the sense of § 3 in that it maps the set of multiple points of u onto the set of multiple points of v, preserving type, and maps $M(u) \mid \beta D_1$ onto $M(v) \mid \beta D_1$ preserving sense on βD_1.

The indexing of stars $S(P, u)$ and $S(Q, v)$. If P and Q are multiple points of u and v respectively with $T_1(P) = Q$ then P and Q are multiple points of U and V with $T(P) = Q$. The sectors of $S(P, u)$ and $S(Q, v)$ will be indexed in such a manner that a sector $S_i(P, u)$ has the index i in common with a sector $S_i(Q, v)$ if and only if the sectors in $S(P, U)$ and $S(Q, V)$ containing $S_i(P, u)$ and $S_i(Q, v)$ respectively as point sets, bear equal indices k.

Finally $M(u) \sim M(v)$ under T_1. The analysis follows. Each multiple point of u or v respectively is a multiple point of U or V. Among points in $\bar{D}_1$, the converse is true except at most for the points $z = \pm 1$. The stars of u [or v] are thus the intersections with $\bar{D}_1$ of stars of U [or V] with the same centers. Note that $M(u) = [M(u) \cap M(U)] \cup d$. Hence $M(u) = [M(u) \cap M(U)] \cup A$, where A is the set of points $z_0 \varepsilon d$ not multiple points of U.

The incidence relations. It follows from the definition of T_1 and the indexing of the sectors of the stars $S(P, u)$ and $S(Q, v)$ that a point

$z^* \varepsilon M(u) \cap M(U)$ has the same incidence relations with closed sectors of $S(P, u)$ as $T(z^*)$ with the corresponding closed sectors of $S(Q, v)$. These points z^* include the multiple points of u in d (in particular possibly $P_0 = Q_0$). These multiple points z^* of u in d are also multiple points of U, V and v with $T(z^*) = T_1(z^*) = z^*$.

It remains to consider a point $z_0 \varepsilon A$. Since z_0 is not a multiple point of U no ray of a star of U or V meets z_0 other than a ray on d. The point z_0 is immediately preceded (or followed) on d by a multiple point z^* of U, V. If z^* is on a sector $S_i(P, u)$ with P not in d, or on a left (or right) boundary ray of $S_i(P, u)$ then z_0 is in $S_i(P, u)$. It follows that z^* is in $S_i(Q, v)$ or on the left (or right) boundary ray of $S_i(Q, v)$. Hence z_0 is in $S_i(Q, v)$. Thus z_0 has the same incidence relations with the closed sectors of $S(P, u)$ as with the corresponding closed sectors of $S(Q, v)$. If $P = Q$ is in d with x-coordinate $a < 1$ then the subarc of d on which $a \leqq x \leqq 1$ is a right boundary arc of the first sector both of $S(P, u)$ and $S(Q, v)$ so that any point x_0 on this arc has the same incidence relations with the closed sectors of $S(P, u)$ as with the closed sectors of $S(Q, v)$. The arc $-1 \leqq x \leqq a$ is similarly treated if $a > -1$. Thus $M(u) \sim M(v)$ under T_1.

The induction. Suppose that $\nu(U) = \nu(V) = n > 0$. Let $\nu(u) = \nu(v) = m$. The inductive hypothesis of the lemma can be applied to u and v if $m < n$. That $m < n$ follows from the fact that each multiple point P of u and v is a multiple point of U and V of no less index, while each multiple point of U and V on d, in particular P_0, has an index relative to u and v which is less than its index relative to U and V. Indeed P_0 may be ordinary relative to u and v. By virtue of the inductive hypothesis u is C. E. with v under a mapping ϕ which is an extension over $\bar{D}_1$ of T_1 over $M(u)$. Thus $\phi(z) = z$ for $z \varepsilon d$, and $\phi(z) = T(z)$ for $z \varepsilon M(U) \cap D_1$. Let D_2 be the semi-disc of D on which $y < 0$ and set $u' = U \mid \bar{D}_2$, $v' = V \mid \bar{D}_2$. As just shown for u and v, u' and v' are C. E. under a mapping ϕ' such that $\phi'(z) = z$ for $z \varepsilon d$ and $\phi'(z) = T(z)$ for $z \varepsilon M(U) \cap \bar{D}_2$. Let a mapping Φ of $\bar{D}$ onto itself be defined by combining the mappings ϕ and ϕ'. Then U and V are C. E. under Φ, and Φ is an extension of T.

This completes the proof of the lemma.

The inductive proof of Th. 4. 1 will be completed by proving the following. This is the case $\nu(U) = \nu(V) = 0$.

LEMMA 6. 2. *Suppose that $U \varepsilon [D]'$ and V in $[\Delta]'$ have no multiple points and that T is a topological mapping of the maximal connected U-arcs*

in βD *onto the maximal connected* V*-arcs in* $\beta\Delta$ *preserving sense on the boundaries.*[6] *There then exists an extension* ϕ *of* T *over* D *under which* U *is* C. E. *to* V.

We begin by verifying the following.

(α). *The only maximal connected* U*-arcs in* βD *are those on which* U *assumes a proper extremum.*

If h were a maximal U-arc in βD on which U was not a proper relative extremum there would be some U-arc in D which would have an end point $P \varepsilon h$. Such a point P would be a multiple point of U coming under Cases V or VII of § 4, contrary to the assumption that U has no multiple points.

(β). *The extreme values of* U *reduce to an absolute minimum and absolute maximum assumed respectively in just one maximal connected* U*-set* p *and one maximal connected* U*-set* q*, where* p *and* q *may be* U*-arcs or points.*

Suppose that there were at least two disjoint maximal connected U-sets affording relative minima. Then for a suitable choice of c the set U_c on which $U(z) \leqq c$ would not be connected. Since the set U_c is connected for $c = \max U(z)$ there exists a superior limit c_0 of the values of c for which U_c is not connected. One proves easily as in M., § 10 that there must be a multiple point P at the U-level c_0. The lemma follows.

Reduced μ*-length* [M., § 27]. If a and b are constants with $a \neq 0$, $aU + b$ is C. E. with U under the identity. No generality is accordingly lost in proving the lemma if we assume that the range of values of U and V is the interval $[0, 1]$. No maximal level set of U or V will then reduce to a point with the possible exception of sets at the level 0 or 1. Each U-arc λ will be referred to its reduced μ-length ρ as parameter. This is the μ-length of λ measured in λ's U-positive sense from λ's initial point on βD, and divided by the total μ-length of λ. On each such λ, ρ varies from 0 to 1 inclusive. The V-arcs will be similarly referred to their reduced μ-lengths σ as parameter.

To begin the proof proper consider first the case in which p and q are maximal connected U-arcs on which U assumes its absolute minimum and maximum respectively, and let $T(p) = p'$ and $T(q) = q'$. The arcs p' and q' may be maximizing and minimizing respectively rather than minimizing and maximizing. No generality in the proof will be lost if we assume that p' and q'

[6] One gives the boundaries a counter-clockwise sense as a basis of reference.

are respectively minimizing and maximizing since a change from V to the C. E. — V would bring this about in any case.

Suppose that the given mapping T of p onto p' has the form $\sigma = \eta(\rho)$, $0 \leqq \rho \leqq 1$, making the point on p with reduced μ-length ρ correspond under T to the point on p' with reduced μ-length σ. The senses of increasing μ-length on p and p' have been taken as the U-positive and V-positive senses on p and p' and so are counter-clockwise on βD and $\beta\Delta$ respectively. From this and the nature of T as given, it follows that $\eta(\rho)$ is *increasing*. Similarly suppose that the given mapping T of q and q' has the form $\sigma = \zeta(\rho)$, $0 \leqq \rho \leqq 1$ and verify the essential fact that $\zeta(\rho)$ is increasing. An admissible extension ϕ of T mapping $\bar{D}$ onto $\bar{\Delta}$ is obtained by making each U-arc λ_c at the level c correspond to the V-arc θ_c at this level c, and making the point ρ on λ_c correspond to the point σ on θ_c such that

$$\sigma = (1-c)\zeta(\rho) + c\eta(\rho) \qquad (0 \leqq \rho \leqq 1). \tag{6.1}$$

Recalling that $\zeta(\rho)$ and $\eta(\rho)$ are both increasing it follows from the properties of reduced μ-length that ϕ extends T as a sense preserving homeomorphism and defines a contour equivalence of U with V. One first verifies that T is a homeomorphism. That T is sense preserving then follows from the fact that it is sense preserving in the neighborhood of one point, in particular in the neighborhood of a point on p.

In case p reduces to a point but q does not, let $T(q) = q'$, and as before suppose V assumes its absolute maximum on q'. Then the maximal connected V-set in which V assumes its absolute minimum must reduce to a point which we denote by p'. Now define reduced μ-lengths on the level arcs and as in the preceding paragraph let the given mapping T of q onto q' have the form $\sigma = \zeta(\rho)$. Then the mapping $\sigma = \zeta(\rho)$ $(0 < c \leqq 1)$ yields the extension ϕ of T provided we require $\phi(p) = p'$.

In case both the minimizing set p and maximizing set q for U reduce to points, the maximal connected V-sets p' and q' in which V assumes its absolute minimum and maximum also reduce to points. Defining reduced μ-length as before, the mapping $\sigma = \rho$ for $0 < c < 1$ together with $\phi(p) = p'$, $\phi(q) = q'$ provide a mapping ϕ which is a contour equivalence of U with V.

This completes the proof of Th. 4.1. Th. 4.1 implies Th. 3.1 as a special case.

7. A group of operators. Given $U \varepsilon [D]$ let $[U]$ be the class of functions in $[D]$ C. E. with U. In this section we shall describe a multiplicative group of operations which generates $[U]$ from U. Certain preliminary remarks are needed.

The sum of two pseudoharmonic functions u and v defined over D is not in general pseudoharmonic over D. For example, set $z = x + iy$ and let

$$u(z) = x, \qquad v(z) = \tfrac{1}{4}x^2 - x \qquad (|z| < 1).$$

The functions u and v have no critical point on the domain $|z| < 1$ and so are pseudoharmonic. However $u + v$ has a minimum when $z = 0$ and so is not pseudoharmonic. The functions u and v in this example are easily seen to be C. E. under the identity. Thus the sum of two functions C. E. under the identity need not be pseudoharmonic. There is nevertheless a law in the background.

Suppose that u and v are in $[U]$ and C. E. under ϕ. Let g be any simple sensed arc in $\bar{D}$ on which u is strictly increasing. Then for each choice of such a g, v will be strictly increasing on $\phi(g)$, or else strictly decreasing independently of the choice of such a g. This may be verified first for arcs g which intersect the set of multiple points of u in at most an end point. For any one such arc g' can be continuously deformed into any other such arc g'' through admissible arcs g. Finally an admissible arc g which intersects the set of multiple points is a sequence of a finite set of admissible arcs g_i each one of which intersects the set of multiple points in at most an end point.

If g is admissible in the above sense and v is strictly increasing on $\phi(g)$, u and v will be termed *positively* C. E. under ϕ, otherwise *negatively* C. E. under ϕ. Whether u and v are positively or negatively C. E. under ϕ is thus determined by the behavior of v on $\phi(g)$ for one arbitrarily chosen admissible arc g. The following are readily verified:

(α). *If u and v in $[U]$ are* C. E. *under ϕ, then either v or else $-v$ is positively C. E. to u under ϕ.*

(β). *The sum of $u \varepsilon [U]$ and $v \varepsilon [U]$, with u and v given as positively* C. E. *under the identity, is again in $[U]$ and positively* C. E. *to u and to v under the identity.*

To generate $[U]$ from U we shall make use of a deformation Δ of elements $u \varepsilon [U]$. The deformation of u is defined by a 1-parameter family of functions in $[U]$ which for fixed t, $0 \leqq t \leqq 1$, and fixed $u \varepsilon [U]$ have values $\Delta(z, t, u)$ $(z \varepsilon \bar{D})$. For fixed u, Δ is supposed continuous over the cartesian product of the domains of z and t. One supposes that u is given "initially" in the form $u(z) = \Delta(z, 0, u)$. The "terminal" image Δu of u under Δ has by definition the values $(\Delta u)(z) = \Delta(z, 1, u)$ $(z \varepsilon \bar{D})$. For fixed $t \varepsilon [0, 1]$

and $u \varepsilon [U]$, $\Delta(z, t, u)$ shall define a function in $[U]$, C. E. with u under the identity.

With this understood we introduce three operations on elements in $[U]$.

(i) *A homotopy* Δ, *replacing* $u \varepsilon [U]$ *by its terminal image* Δu *under* Δ.

(ii) *A reflection* R, *replacing* u *by* $Ru = -u$.

(iii) *A value-equivalence* Φ, *replacing* u *by* $u\phi$ *where* ϕ *is a* S. P. *homeomorphism mapping* $\bar{D}$ *onto* $\bar{D}$.

Each of these operations on u yields an image C. E. with u. In cases (i) and (ii), Ru, and Δu are C. E. with u under the identity. Each of these operations has an inverse. The inverse of R is R. The operation Φ determined by the mapping ϕ has an inverse Φ^{-1} determined by ϕ^{-1}. If Δ is defined by $\Delta(z, t, u)$ as above, the inverse of Δ is defined by $\Delta(z, 1-t, u)$ for $u \varepsilon [U]$, $z \varepsilon \bar{D}$, and $0 \leqq t \leqq 1$. The operations R, Φ, Δ generate a group Ω.

The principal theorem can now be stated.

THEOREM 7. 1. *Each element* $u \varepsilon [U]$ *has the form* ωU *where* ω *is an element in the group* Ω.

Given $u \varepsilon [U]$, U is C. E. with u under some mapping ϕ. Hence $u\phi$ is C. E. with U under the identity. In accordance with (α), $R_0 U$ is positively C. E. with $u\phi$, where $R_0 = R$, or the identity. We set $u\phi = \Phi u$ and introduce the deformation

$$\Delta(z, t, R_0 U) = (1-t)(R_0 U)(z) + t(\Phi u)(z) \qquad (0 \leqq t \leqq 1). \tag{7.1}$$

For $0 < t < 1$ the two terms in the right member define positively C. E. functions whose sum is in $[U]$ by virtue of (β). As a consequence of (7. 1)

$$\Delta R_0 U = \Phi u \rightarrow \Phi^{-1} \Delta R_0 U = u \tag{7.2}$$

This establishes the theorem.

The first relation in (7. 2) has the following meaning.

COROLLARY. *If* $u \varepsilon [D]$ *and* $U \varepsilon [D]$ *are* C. E. *under the identity there exists a continuous deformation of* u *through elements in* $[U]$, *each* C. E. *with* U *under the identity, into one of the two elements* $\pm U$.

Part II. Pseudoconjugates.

8. Definition. Let u be pseudoharmonic in D. A pseudoharmonic function v such that $u + iv$ is interior in D will be called pseudoconjugate

to u in D. If u is continuous on $\bar{D}$ and pseudoharmonic in D, a function v continuous in $\bar{D}$ and pseudoconjugate to u in D will be termed pseudoconjugate to u *in* $\bar{D}$. When v is pseudoconjugate to u in D a point $z_0 \varepsilon D$ is an ordinary or multiple point of u of index n, if and only if it has the same character relative to v. We shall simplify the problem of constructing pseudoconjugates by noting the following:

(β). *If v is pseudoconjugate to $u \varepsilon [D]$, and if ϕ is any* S. P. *homeomorphism of $\bar{D}$ onto itself, then $v\phi$ is pseudoconjugate to $u\phi$.*

This follows at once from the definition of a pseudoconjugate.

As a consequence of (β) to construct a pseudoconjugate to $U \varepsilon [D]$ one can use any domain E such that $\bar{E}$ is the topological image $\phi(\bar{D})$ of $\bar{D}$ under a S. P. homeomorphism ϕ, and replace U by $u \varepsilon [E]$ where $U = u\phi$. If then v is constructed pseudoconjugate to u, $v\phi$ is pseudoconjugate to U.

(γ). *Let u and v be pseudoharmonic functions in $\bar{D}$ such that $u + iv = f$ maps $\bar{D}$ topologigcally into the complex w-sphere. Then f is sense-preserving and v therefore pseudoconjugate to u if there exists a continuous* 1-*parameter family of topological mappings f^t $(0 \leqq t \leqq 1)$ of $\bar{D}$ onto the complex w-sphere such that $f^0 = f$ and f^1 is interior.*

This follows from the primitive definition of a S. P. topological mapping. In applying (γ) one can take f^1 as analytic, or if convenient as the identity.

9. Three special constructions of pseudoconjugates. By virtue of the remarks of § 8 the three special constructions of pseudoconjugates now to be given have wide application. We refer to the complex plane of $z = x + iy$.

I. *Let E_1 be the square $0 \leqq x \leqq 1$, $0 \leqq y \leqq 1$. On E_1 let u be the pseudoharmonic function with the values $ax + b$, $a > 0$. Let $h(y)$, $0 \leq y \leq 1$, be continuous and strictly increasing. Then a function v with values*[7] *$v(z) = h(y) + x$ is pseudoconjugate to u.*

Since $u + iv$ clearly defines a topological mapping of the z-plane into another complex plane the only point of difficulty is in proving that $u + iv$ is sense preserving. In accordance with (γ) of § 8 this is established by deforming $h(y)$ through a continuous 1-parameter family of strictly increasing functions into the identity. When $v(z) = y + x$, $u + iv$ is clearly sense preserving and I follows.

[7] The x is added to make the boundary values of v strictly monotone when $y = 0$, or 1.

II. *Let E_2 be the semi-disc: $x \leqq 0, |z| \leqq 1$ with diameter d on which $x = 0$. Let u be the pseudoharmonic function with values $ax + b, a > 0$. Let $h(y)$ be continuous and strictly increasing. Then a function v with values $v(z) = h(y)$ is a pseudoconjugate of u.*

III. *On the disc $E_3: |z| \leqq 1$ the function v with values y is pseudoconjugate to the function u with values $ax + b$, where $a > 0$.*

The pseudoconjugates v defined in I, II and III lead to more general constructions as follows: Let $R \subset D$ be a Jordan domain such that $\bar{R}$ is mapped by a S. P. homeomorphism ϕ onto $\bar{E}_i$, $i = 1, 2, 3$. Then the functions $U = u\phi$ and $V = v\phi$ are defined over $\bar{R}$ and V is a pseudoconjugate of U. More definitely we limit R to subregions R_i of D as follows:

Case I. $\bar{R}_1$ shall be bounded in $\bar{D}$ by two disjoint arcs of βD and by two simple disjoint non-intersecting arcs p and q, intersecting βD in a finite number of points including the end points of p and q.

Case II. $\bar{R}_2$ shall be bounded in $\bar{D}$ by an arc of βD and by a simple arc d, in D except for a finite number of points, including d's end points.

Case III. R_3 shall be identical with D.

By virtue of the constructions I, II and III functions U pseudoharmonic on R_i for which pseudoconjugates V exist can be characterized together with V as follows (brackets indicate alternative):

Class I. Suppose that a $U \varepsilon [\bar{R}_1]'$ (See § 4) assumes its minimum (maximum) and its maximum (minimum) respectively at each point of the arcs p, q of βR_1, is strictly monotone on the two complementary arcs of βR_1 and has no multiple points at any point of $\bar{R}_1$. A pseudoconjugate V of such a U always exists with values which are presecribed on p and strictly decreasing [8] (increasing), which has no multiple points in $\bar{R}_1$ and is strictly monotone on βR_1 except for a minimum (maximum) and maximum (minimum) at the final points of p and q respectively.

Class II. Let a $U \varepsilon [R_2]'$ assume its absolute maximum (minimum) at each point of the arc $d \varepsilon \beta R_2$, assume its absolute minimum (maximum) at a point $P \varepsilon \beta R_2$, possess no other extreme boundary points and no multiple points. A pseudoconjugate V of U then exists without multiple points in $\bar{R}_2$,

[8] Sense of increase of functions on βR_i are with respect to an independent variable moving on βR_i in a counter-clockwise sense. The arcs p and q are to be sensed counter-clockwise on βR_1 for this purpose.

with prescribed strictly increasing (decreasing) values in d and strictly decreasing (increasing) values in $\beta R_2 - d$.

Class III. Let $U \varepsilon [D]$ have no multiple points and $U \mid \beta D$ just two extreme points. Then there exists a pseudoconjugate V of U of the same character as U such that the extreme points of U and V appear in βD in the circular order min U, min V, max U, max V.

In establishing the existence of the pseudoconjugates V one begins by showing that the respective functions U in Classes I, II, III are strictly C. E. with $\pm u$ in I, II, III, making use of μ-lengths along the U-positive sensed level arcs (Cf. §§ 5, 6) to obtain the appropriate mapping ϕ. The existence of pseudoconjugates to functions U in Classes I, II, III is all that we shall need to establish the existence of a pseudoconjugate to an arbitrary $u \varepsilon [D]$.

10. Secteurs and inner boundaries. A pseudoconjugate of an arbitrary $U \varepsilon [D]$ will eventually be constructed out of the special constructions of § 9. To that end D must be broken up into special regions X coming under Cases I, II, and III, with $U \mid X$ in Classes I, II, III respectively. We shall need several definitions.

If g is a sensed arc the corresponding unsensed arc will be denoted by $\mid g \mid$. We say that $\mid g \mid$ *carries* g.

The net $\mid N(U) \mid$. Let z_0 be an arbitrary multiple point of U. Let h be a U-arc issuing from z_0 and continued in the ordinary sense until βD or another multiple point is reached. Let $N(U)$ be the union of all such sensed U-arcs. Let $\mid N(U) \mid$ be the union of the corresponding unsensed arcs.

Right or left continuations. An arc $h_2 \varepsilon N(U)$ is termed the ***right*** (***left***) continuation of an arc $h_1 \varepsilon N(U)$ if h_1 terminates in the initial point P of h_2 and if $\mid h_1 \mid$ and $\mid h_2 \mid$ are the right and left rays respectively (left and right rays) as seen from P, bounding a sector in $S(P, U)$. A simple sensed arc $h'(h'')$ composed of a sequence of arcs of $N(U)$ is termed the ***maximal right*** (***left***) ***continuation*** of each of its subarcs in $N(U)$ if the second of any two successive arcs in $N(U)$ and $h'(h'')$ is the right (left) continuation of the first, and if $h'(h'')$ is a proper subarc of no arc with this property. These maximal right (left) continuations are obviously simple U-arcs with end points on βD.

Inner boundaries. The preceding maximal right (left) continuations $h'(h'')$ if reversed in sense are maximal left (right) continuations. An

unsensed arc k identical with $|h'|$ or $|h''|$ will separate D into two or more regions the closure of one and only one of which (termed a *secteur* K) will contain k and no elements of $|N(U)|$ incident with k other than elements of $|N(U)| \cap k$. Observe that $D - \bar{K}$ is not a secteur. The secteur K is bounded by k and a unique arc of βD. We term k an *inner boundary.* It is clear that K determines and is determined by its inner boundary k. Two inner boundaries either do not intersect or intersect in a point or arc of $|N(U)|$. If k' is a second inner boundary and if $k \cap k' \neq 0$, k' is in $D - K$. Of the regions into which k separates D, the secteur K determined by k is distinguished from the other components of $D - k$ by the following property: if λ and μ are two elementary arcs of $k \cap |N(U)|$ with a multiple point P as common end point, the sector of a canonical neighborhood of P bounded by λ and μ belongs to K. Every element $|h_1|$ in $|N(U)|$ belongs to just two inner boundaries, carrying the maximal right and left continuations of h_1.

Let z be a point not in $|N(U)|$. Of the connected regions into which $|N(U)|$ divides D let X be the region containing z. If βX includes an arc $|h_1| \in |N(U)|$, it is clear that it must include the carrier p of either the right or left continuation of h_1. The U-arc p is the inner boundary of some secteur K. Then either $X \subset K$ or else $X \cap K = 0$. The case $X \subset K$ prevails for the following reasons. If P is a multiple point of U in p then X intersects $S(P, U)$ near P in just one sector S' of $S(P, U)$ and p contains both arcs of $|N(U)|$ on the boundary of S' issuing from P. It is characteristic of $\bar{K}$ that, containing both of these arcs of $|N(U)|$, it contains the sector S' in a neighborhood of p. Thus $X \subset K$.

LEMMA 10.1. *The net $|N(U)|$ (assumed non-empty) divides D into a finite set of connected open regions X containing no multiple points of U. Each region X is either,*

(i) *the intersection R of two secteurs K and G possessing non-intersecting inner boundaries p and q, with R bounded by p, q and two disjoint arcs of βD; or*

(ii) *a secteur K.*

In Case (i), $U \mid \bar{R}$ *is in Class* I *of* § 9. *In Case* (ii), $U \mid \bar{K}$ *is in Class* II.

If X is bounded by p and an arc of βD, then X is a secteur K and Case (ii) alone occurs. In this case it is clear that $U \mid \bar{K}$ is in Class II of § 9.

If Case (ii) does not arise, βX contains a second inner boundary q, the inner boundary of some secteur G. As before $G \supset X$. Moreover $p \cap q = 0$;

otherwise q would include arcs of $|N(U)|$ in the complement of $\bar{K}$, which is impossible since $\bar{X} \subset \bar{K}$. We see that βX includes at most a finite set of disjoint maximal connected U-arcs. But $U \mid \bar{X}$ has no multiple points, so that it follows as in § 6 (α) and (β) that βX contains just two disjoint inner boundaries p and q, and $U \mid \bar{X}$ is in Class I of § 9.

11. The general construction of pseudoconjugates. If $U \varepsilon [D]$ has no multiple points at all, U is in Class III of § 9 (Cf. (β) of § 6), and a pseudoconjugate V of U is imemdiate. We assume therefore that U has at least one multiple point.

It will be convenient to term the region R arising in Case (i) of Lemma 10.1 a *secteur band.* By means of the net $|N(U)|$, $\bar{D}$ has been decomposed into a finite number of secteur bands and secteurs on which U is in Class I and Class II respectively. To construct pseudoconjugates of U over the closure of any one of these regions X so as to yield a resultant continuous pseudoconjugate V of U one must progressively assign boundary values of V (termed V-values) along preferred U-arcs of βX as in the constructions of § 9.

V-values. If z_1 is an ordinary point of U on a U-arc g and if g is sensed U-positively [U-negatively] (see § 5) then a pseudoconjugate V of U *must* decrease [increase] along g near z_1 in order that $U + iV$ may be sense preserving.

To construct a function V pseudoconjugate to U we prefer an arbitrary one of the secteurs K or secteur bands R into which D is separated by $|N(U)|$.

Case a. In the case of K, continuous strictly monotone values (termed V-values) will be arbitrarily assigned along the inner boundary k of K so that these values increase or decrease in the unique sense possible for a pseudoconjugate of U. Recall that $U \mid \bar{K}$ is in Class II of § 9. (Lemma 10.1.) Hence U possesses a pseudoconjugate V over $\bar{K}$ extending the V-values just assigned along k. The resulting $V \mid \bar{K}$ is strictly monotone over $\beta D \cap \beta K$.

Case b. If the preferred region is a secteur band R we choose one of the inner boundaries k in βR and assign V-values along k as in Case a. In accordance with Lemma 10.1, $U \mid \bar{R}$ is in Class I of § 9. There accordingly exists a pseudoconjugate V of U over $\bar{R}$ extending the V-values assigned along k. Along the second inner boundary k' in βR, V, as constructed over $\bar{R}$, gives strictly monotone values, increasing or decreasing in the unique manner possible for a pseudoconjugate of $U \mid \bar{R}$.

Let X be any secteur or secteur band in the decomposition of D by $|N(U)|$ such that βX intersects βK, Case a, or βR, Case b, in an arc. This intersection must be along a single connected arc g, since no U-arc can be closed in $\bar{D}$. Let Σ denote K or R according as Case a or Case b arose in the first construction. Let $k \supset g$ be the inner boundary of X which intersects $\beta\Sigma$. Note that $g \neq k$. Recall that V-values have already been constructed on g. V-values will now be assigned continuously on the residual arc of k so that the V-values are strictly monotone over all of k. A function V pseudoharmonic over X can then be constructed as in Case a or Case b extending the values given on k.

In general let Σ be the closure of the union of the secteurs and secteur bands on which V has already been constructed. As in the second stage of the process let X represent any secteur or secteur band in the decomposition of D such that $\beta X \cap \beta\Sigma$ is an arc g. Let $k \supset g$ be the inner boundary of X which intersects $\beta\Sigma$. In the general case it is possible that $k = g$. In case $k = g$, V-values have already been constructed along k. In case $k \neq g$, V-values are assigned along k as in the second stage. A function V is then constructed over X as in the second stage. This process is continued until V is constructed over all of $\bar{D}$.

It may be remarked that the construction of V is such that $V \mid \beta D$ is strictly monotone over each of the arcs of βD into which βD is separated by $|N(U)|$.

It remains to prove that $f = U + iV$ is interior over D.

It is interior by construction at each point $z_0 \varepsilon D$ not in the net $|N(U)|$. Let z_0 then be in $|N(U)| \cap D$ *but* not a multiple point of U. Without loss of generality we can suppose that in a sufficiently small neighborhood of z_0, $U(z) = x + c$, for this would be true of a composite function $U\phi$ where ϕ was a suitable S. P. homeomorphism of a neighborhood of z_0. As explicitly constructed V is continuous and strictly decreasing along each U-arc (with U-positive sense) sufficiently near z_0, so that it is clear that f is sense preserving and topological in a sufficiently small neighborhood of z_0.

Suppose finally that $z_0 \varepsilon D$ is a multiple point of U. If $e > 0$ is sufficiently small and N_e is the neighborhood $(|z - z_0| < e)$ of z_0, $f \mid N_e$ satisfies the conditions of Lemma 2.1 and is accordingly interior at z_0. The following theorem is accordingly proved.

THEOREM 11.1. *Corresponding to an arbitrary $U \varepsilon [D]$ there exists a pseudoconjugate $V \varepsilon [D]$ of U pseudoharmonic over D and continuous over $\bar{D}$.*

12. The continuation of interior transformations. To complete our results the following theorem is needed:

THEOREM 12. 1. *An interior mapping f of a Jordan domain D into the w-plane, continuous on $\bar{D}$, which send βD into a locally simple curve g* [M., p. 62] *can be extended to an interior transformation over a Jordan domain $\Delta \supset \bar{D}$.*

To establish this theorem a Riemann ribbon Σ spread over the w-plane and bearing a curve g' which is simple on Σ and projects into g will be constructed. The more precise statement is as follows:

LEMMA 12. 1. *Given a locally simple sensed curve g in the w-plane there exists an interior mapping Z of the form $w = Z(z)$ of a domain*

$$a < |z| < a' \qquad (0 < a < 1) \qquad (aa' = 1) \tag{12.1}$$

of the z-plane into the w-plane which sends the unit circle $(|z| = 1)$ in counter-clockwise sense into g, and which is such that for some constant $e > 0$ and for each subdomain of (12. 1) for which $\theta_0 - e \leqq \arg z \leqq \theta_0 + e$ the mapping Z is topological.

The Riemann surface of Z^{-1} is the Riemann ribbon to which reference was made.

To prove this lemma let $\cdots w_{-2}, w_{-1}, w_0, w_1, w_2, \cdots$ be a cyclic sequence of points on g with $w_i = w_{i+n}$ for some $n > 3$ and all i, so chosen that not only the subarcs $g_i = [w_i, w_{i+1}]$ but also the subarcs

$$[w_i, w_{i+1}, w_{i+2}, w_{i+3}] \tag{12.2}$$

of g are simple. Let k_i $(i = 0, \pm 1, \pm 2, \cdots)$ be a simple Jordan arc in the w-plane such that k_i intersects the subarc $[w_{i-1}, w_i, w_{i+1}]$ of g in w_i alone, where w_i is an interior point of k_i and where $k_i = k_{i+n}$ for all i. Because of the simplicity of the arc (12. 2) we may assume that k_i intersects neither k_{i-1} nor k_{i+1}. There exists a sense-preserving topological mapping T_r of the z'-plane into the w-plane which sends the arcs $[y' = 0, 0 \leqq x' \leqq 1]$ into g_r and the arcs $[x' = 0, -1 \leqq y' \leqq 1]$ and $[x' = 1, -1 \leqq y' \leqq 1]$ into k_r and k_{r+1} respectively. The mapping T_r parametrizes k_r and k_{r+1} in terms of y'. We can suppose $T_0, T_{\pm 1}, T_{\pm 2}, \cdots$ successively chosen so that T_r and T_{r+1} give the same parameterization to k_{r+1}, and so that $T_i = T_{i+n}$ for all i. In particular T_{n-1} can be chosen so as to give the same parameterization to $k_0 = k_n$ as does T_0, since the w-plane is orientable. The condition of simplicity for the arcs (12. 2) has the following consequence. If $b > 0$ is sufficiently

small, then for z' and z'' in the rectangle $H: 0 \leqq x' \leqq 1,\ -b \leqq y' \leqq b$, $T_r(z') = T_{r+1}(z'')$ if and only if $z' - z'' = 1$ and if z'' is pure imaginary or null. That is for z' and z'' in H, the mappings T_r and T_{r+1} coincide only in their mapping of $-b \leqq y' \leqq b$ onto k_{r+1}.

Let T now map the strip $[-\infty < x' < \infty, -b \leqq y' \leqq b]$ into the w-plane with T defined by setting

$$T(z' + r) = T_r(z') \qquad (r = 0, \pm 1, \pm 2, \cdots)\ (z' \varepsilon H). \tag{12.3}$$

Under T the map of the x'-axis is an unending locally simple sensed curve which reduces to g by virtue of the relation $T(x' + n) = T(x')$. If e_1 is a sufficiently small positive constant and if c is an arbitrary value of x' the mapping T taken over the rectangle

$$c - e_1 \leqq x' \leqq c + e_1, -b \leqq y' \leqq b, \tag{12.4}$$

is sense-preserving and topological. For z' in the domain of definition of T set $z = \exp(2\pi i z'/n)$ and subject to this relation set $Z(z) = T(z')$. The resulting mapping Z has the properties affirmed in the lemma.

Proof of Th. 12. 1. It follows from Lemma 2. 2 that f can be extended as an interior transformation over the neighborhood of each point z_0 of βD neighboring which in $\bar{D}$, f fails to be topological. Since there is at most a finite set of such points on βD [M., p. 85] one can suppose D replaced by a Jordan domain $\Delta \supset D$, over whose closure $\bar{\Delta}$, f can be extended so as to be interior in Δ, continuous in $\bar{\Delta}$ and topological in some neighborhood (relative to $\bar{\Delta}$) of each point of $\beta\Delta$. Without loss of generality we can finally suppose that f is interior over a unit disc D: $(|z| < 1)$, continuous on $\bar{D}$ and locally topological in some neighborhood (relative to $\bar{D}$) of each point of βD.

Such an f maps some neighborhood N (relative to $\bar{D}$) of each point z_0 of βD in a topological manner onto a subset of a neighborhood of the point $f(z_0)$ of g. The points in $f(N)$ are either in g or locally on one side of g. Turning to Lemma 12. 1 let Σ_a be the Riemann surface of Z^{-1} over the w-plane, understanding that $Z(z)$ is defined only over (12. 1). Under Z, βD goes into a curve g' on Σ_a where g' is simple on Σ_a and projects onto g. If the constant $b > 0$ is sufficiently small f maps the domain E_b: $[1 - b \leqq |z| \leqq 1]$ in a topological manner into the closure of one of the two sets, say G_a, into which Σ_a is separated by g'. We suppose b so restricted.

Understanding that $f(z)$ is a point in $\bar{G}_a$ and not merely a point in the w-plane it is clear that $Z^{-1}f(z) = \Phi(z)$ is uniquely defined for $z \,\varepsilon\, E_b$ and that Φ maps E_b topologically into the z-plane, thereby mapping the circle

($|z| = 1$) onto itself topologically. We extend Φ over the reflection E'_b of E_b in the unit circle by setting $\Phi(z) = (\Phi(z'))'$ where the primes denote reflection in the unit circle. Noting that $Z\Phi(z) = f(z)$ for $z \varepsilon E_b$, we extend f over E'_b by requiring this relation to hold over E'_b. Since Z is locally topological over E'_b, f is likewise locally topological over E'_b and hence interior over $E_b \cup E'_b$. The proof of Theorem 12.1 is complete.

13. Existence of a harmonic function contour equivalent to a given $u \varepsilon [D]$. The present section is devoted to the proof of the following result.

THEOREM 13.1. *If D is a Jordan domain and u a pseudoharmonic function in $[D]$ there exists a harmonic function U in $[D]$ such that U is strictly* C. E. *to u.*

In order to apply the results of § 12 we introduce the following lemma.

LEMMA 13.1. *If D is a Jordan domain and $u \varepsilon [D]$ there exists a Jordan domain $\Delta \supset D$ and an extension u^* of u with $u^* \varepsilon [\Delta]$ such that u^* has no multiple points in $\beta\Delta$.*

Let $a \varepsilon \beta D$ be a multiple point of u to which, as a limiting point, m u-arcs tend. Without essential loss of generality suppose that $u(a) = 0$. In the w-plane consider the harmonic function H with values $H(w) = \mathcal{R} w^m$. ($\mathcal{R}$ denotes real part.) Let W be the neighborhood of $w = 0$ in which $|w| < e$. Let W be divided into the semi-disc W': $|w| < e$, $0 < \arg w < \pi$ and the complement of the latter relative to W. Let $\mathfrak{D}$ be the semi-disc defined by $|w| < r$, $0 < \arg w < \pi$ $(r > e)$. The function $H \mid \bar{W}'$ has m level arcs tending to $w = 0$ in W'. Moreover $H \mid \bar{W}'$ is strictly C. E. to $\pm u$ over a canonical neighborhood of $z = a$ in $\bar{D}$. This contour equivalence can be extended as a S. P. homeomorphism ψ of the w-plane onto the z-plane, mapping $w = 0$ onto $z = a$, $\mathfrak{D}$ onto D and, $\beta\mathfrak{D}$ onto βD and such that

$$u(\psi(w)) = \sigma H(w) \qquad (w \varepsilon \bar{W}', \sigma = \pm 1),$$

provided e is sufficiently small.

We extend u as a pseudoharmonic function across βD near $z = a$ by setting $u(\psi(w)) = \sigma H(w)$ $(w \varepsilon W)$. Now consider a Jordan domain D' obtained from D by a modification of βD on a short arc containing $z = a$ in its interior. This modification corresponds under ψ to a change conveniently made in the w-plane, recalling that the segment g of the real w-axis from $w = -e$ to $w = e$ is sent by ψ into an arc of βD containing $z = a$. Let the arc of g from $w = -e/2$ to $w = e/2$ be replaced by a circular arc γ on

$|w| = e/2$, lying in the lower half w-plane apart from its end points. If g' is the open arc thereby replacing g, $H \mid g'$ has $m-1$ extreme points, on γ, not multiple points of $H \mid W$. Through every other point of g' just one level arc of H enters the domain bounded by g' and the arc $|w| = e$, $0 \leq \arg w \leq \pi$. On the arc corresponding to g' under ψ in the z-plane there is thus no multiple point of u. Making corresponding extensions of u and modifications of the boundary in a neighborhood of every other multiple point of u in βD, we obtain the desired domain Δ. The extension u^* of u to Δ is clearly in $[\Delta]$ and has no multiple point in $\beta\Delta$.

Now let us suppose that u in Th. 13.1 has been extended to the Jordan domain Δ so as to have these properties. Let v be a pseudoconjugate of u in $\bar{\Delta}$ constructed according to the prescriptions of §§ 9, 11. Then $u \mid \beta\Delta$ and $v \mid \beta\Delta$ have no common extreme point. Indeed, if $|N(u)|$ now denotes the net of u as extended to Δ, the extrema of $v \mid \beta\Delta$ occur at most at points of $|N(u)|$ in $\beta\Delta$, while the extrema of $u \mid \beta\Delta$ occur at most on the open arcs into which $\beta\Delta$ is divided by the points of $|N(u)| \cap \beta\Delta$. Thus $f = u + iv$ is interior at the points of Δ and maps $\beta\Delta$ on a locally simple curve.

By Th. 12.1 f can be extended to a domain Σ containing $\bar{\Delta}$ (and so $\bar{D}$) in its interior. The Riemann surface which is the image of Σ under f can be mapped conformally, say by a function F, into the finite z-plane. Corresponding to D under Ff we obtain a Jordan region E. Next E can be mapped conformally, say by a function G, onto D, and by a well-known result the mapping can be extended to a homeomorphism between $\bar{E}$ and $\bar{D}$. The compound mapping GFf is thus a S. P. homeomorphism ϕ of D onto itself such that $f\phi^{-1}$ is a regular function on D. Thus the function $U = u\phi^{-1}$ is harmonic on D, continuous on $\bar{D}$ and clearly strictly C. E. to u.

It should be remarked that the proof in the last paragraph could also be carried out by first extending f to a domain containing Δ as in the first paragraph of the proof of Th. 12.1 so as to eliminate partial branch elements [M., pp. 83, 85], and then using the above conformal mapping theorem for closed Jordan domains in a neighborhood of each boundary point relative to the closed domain. In this way one would not require the full force of Th. 12.1.

14. Model functions U and nets $|N(U)|$. Suppose that D is the domain $(|z| < 1)$. Recall that the net $|N(U)|$ of U is defined in § 10. We seek a topological model for the net of U under S. P. homeomorphisms $T(D)$ of $\bar{D}$ onto $\bar{D}$. Recall that the domain D may be regarded as a hyper-

bolic plane with circles orthogonal to $(|z|=1)$ as its straight lines. The closure in $\bar{D}$ of any such arc will be called an H-line.

THEOREM 14.1. *There exists a topological model in $\bar{D}$ for the net $|N(U)|$ of $U \varepsilon [D]$ which is the union of a finite set of H-lines in $\bar{D}$.*

Let g be the continuation in D in the sense of § 2 of any one of the elements $h \varepsilon N(U)$. We term $|\bar{g}|$ a *complete arc* in $|N(U)|$. Two points of $\bar{D}$ will be said to be *cofinite* if both are in D or both in βD.

Set $N_0 = |N(U)|$. By a *vertex* of N_0 is meant a multiple point of U or a point of N_0 in βD. By an *element* of N_0 is meant any arc of N_0 whose end points are vertices of N_0 but which carries no other vertices of N_0. Given two vertices P_r, P_s of N_0 let $N_0(P_r, P_s) = \infty$ if P_r, P_s are not connected on N_0. If P_r and P_s are connected on N_0 let $N_0(P_r, P_s)$ be the minimum number of elements of N_0 which is necessary to traverse to pass from P_r to P_s. It is clear that $N_0(P_r, P_s) = N_0(P_s, P_r)$. We term $N_0(P_r, P_s)$ the N_0-*distance* from P_r to P_s.

We suppose that the vertices of N_0

$$P_1, P_2, \cdots, P_n \tag{14.0}$$

have been ordered as follows. Choose P_1 arbitrarily among vertices of N_0. Choose next the vertices $P_i (i > 1)$ for which $N_0(P_1, P_i)$ is finite taking these vertices in the order of magnitude of $N_0(P_1, P_i)$, and arbitrarily when the N_0-distances from P_1 are equal. Suppose that $P_1, P_2, \cdots, P_r$ have been so ordered and that this set includes every vertex connected on N_0 to any member of the set. Choose P_{r+1} arbitrarily among the remaining vertices of N_0 (if there are any). Follow P_{r+1} by the vertices $P_j (j > r+1)$ which are connected to P_{r+1} on N_0 taking these vertices in the order of magnitude of the numbers $N_0(P_{r+1}, P_j)$ or arbitrarily if the N_0-distances from P_{r+1} are equal. This process will suffice to order the vertices of N_0.

There exists a S. P. homeomorphism T_1 of $\bar{D}$ onto $\bar{D}$ with $T_1(P_1) = P_1$ under which each complete arc in N_0 meeting P_1 has an image which is an H-line. This is readily established on using the Jordan separation and Schoenflies mapping theorems.

Proceeding inductively suppose that for $0 < r < n$ there exists a S. P. homeomorphism T of $\bar{D}$ onto $\bar{D}$ such that the image net $T(N_0) = N$ and the image vertices $T(P_i) = Q_i$, $i = 1, \cdots, n$, have the property that each complete arc h of N meeting the set

$$Q_1, \cdots, Q_r \tag{14.1}$$

is an H-line. This has been established for $r=1$. Two cases are to be distinguished.

Case I. Q_{r+1} is connected to Q_r on N.

Case II. Not Case I.

Case I. In this case some point Q_j, $j \leqq r$, and Q_{r+1} determine an H-line b of N. There exists a point Q on b cofinite with Q_{r+1}, with Q_j, Q_{r+1}, Q in the order written on b, and with Q so near βD that there are H-lines meeting Q which do not meet the H-lines h of N except b. (If Q_{r+1} is in βD we take $Q = Q_{r+1}$.) There will then exist a homeomorphism T^1 of $\bar{D}$ onto $\bar{D}$ for which $T^1(Q_1) = Q_1, \cdots, T^1(Q_r) = Q_r$ and under which the H-lines of N meeting the set (14. 1) are arc-wise invariant, while the image under T^1 of each complete arc of N meeting Q_{r+1} is an H-line meeting Q. The above inductive hypothesis made on the sets (14. 0), (14. 1) and T, is now seen to be satisfied by the sets $P_1, \cdots, P_n$, $Q_1, \cdots, Q_r$, Q, and T^1T; in particular $T^1T(P_{r+1}) = T^1(Q_{r+1}) = Q$.

Case II. Let N^1 be the subset of complete arcs of N meeting the set (14. 1). The point Q_{r+1} lies in a secteur or secteur band R of N^1. One replaces each complete arc of N meeting Q_{r+1} by an H-line through Q_{r+1}. Then the inductive hypothesis holds for suitable T and r replaced by $r+1$.

The theorem follows by induction.

We turn to an inverse problem: What are the characteristics of a set of H-lines in $\bar{D}$ that it may serve as the net of some $U \varepsilon [D]$? The answer is in terms of the following definition.

DEFINITION. *We admit any finite set N^* of H-lines in D each of which intersects at least one other H-line and which does not include the entire boundary of any domain in D.*

According to Theorem 14. 1 there exists a topological model of the net $|N(U)|$ of a $U \varepsilon [D]$ which is a set N^* of H-lines as admitted above. Conversely we have the theorem

THEOREM 14. 2. *Corresponding to any admissible set N^* of H-lines there exists a $U \varepsilon [D]$ for which $N^* = |N(U)|$.*

It is clear that N^* separates $\bar{D}$ into a finite number of secteurs or secteur bands R. The proof of the existence of a $U \varepsilon [D]$ for which $N^* = |N(U)|$ is similar to the proof of the existence of a pseudoconjugate $V \varepsilon [D]$ as given in § 11. One adds the regions R successively so as to always keep a simply

connected domain. After the first step, U is given by the previous construction of U as constant on a single arc of $\beta R \cap N^*$. In the case of a secteur R, U is first extended as a constant c over all of $\beta R \cap N^*$ and then extended over R with $U < c$ or $> c$ according as $U > c$ or $< c$ on the secteurs or secteur bands with which βR has arcs in common. No inconsistency can appear as a consequence of this demand at a vertex of N^* in βD nor at any interior vertex P, since there is always an even number of arcs of N^* incident with P. In the case of the adjoining of a secteur band R the value c on one boundary arc of R in N^* is determined by the previous construction of U. The value c' on the other boundary arc of R in N^* is arbitrary subject to one of the conditions $c' < c$, or $c' > c$ uniquely determined by the previous construction.

The theorem follows.

The preceding suggests a weaker form of Th. 4. 1.

THEOREM 14. 3. *A necessary and sufficient condition that u_1 and u_2 in $[D]$ with nets N_1 and N_2 respectively be contour equivalent under some* S. P. *homeomorphism ϕ of $\bar{D}$ onto $\bar{D}$ is that there exist a* S. P. *homeomorphism t of $\bar{D}$ onto $\bar{D}$ under which $t(N_1) = N_2$. If ϕ exists one can take $t = \phi$. If t exists one can take ϕ as an extension of $t \mid N_1$.*

If ϕ exists then $t = \phi$ satisfies the condition of the theorem. If t exists it is clear that in Th. 4. 1, $M(u_1)$ is similar to $M(u_2)$ under $t \mid N_1$ and in accordance with Th. 4. 1 one can take ϕ as an extension of $t \mid N_1$.

THE JOHNS HOPKINS UNIVERSITY AND
THE INSTITUTE FOR ADVANCED STUDY.

BIBLIOGRAPHY.

[1] M. Morse and M. Heins, "Deformation classes of meromorphic functions and their extensions to interior transformations," *Acta Mathematica*, vol. 79 (1947), pp. 51-103.

[2] James A. Jenkins, "On the topological theory of functions," to appear in the *Canadian Journal of Mathematics*.

THE ORTHOGONAL GROUP IN HILBERT SPACE.*

By CALVIN R. PUTNAM and AUREL WINTNER.

1. Let $\mathfrak{R}$ denote the real Hilbert space (in its realization in terms of vectors $x = \{x_k\}$ the components of which are real numbers x_k satisfying $|x| \equiv (\Sigma x_k^2)^{\frac{1}{2}} < \infty$). If A is a real, bounded matrix or, equivalently, a linear (distributive and continuous) operator transforming every point, x, of $\mathfrak{R}$ into a point, Ax, of $\mathfrak{R}$, let $|A|$ denote the least upper bound of the vector length $|Ax|$ when x varies over the unit sphere, $|x| = 1$, of $\mathfrak{R}$.

If an infinite matrix is real and unitary (hence bounded), it will be called an orthogonal matrix, O. Let Ω denote the metric space in which the points are the orthogonal matrices, O, and on which the distance is defined to be $|O_1 - O_2|$ (with the above meaning of $|A|$ for the difference, A, of two orthogonal matrices). Since $|O| = 1$ holds for every O, no distance $|O_1 - O_2|$ in Ω can exceed 2.

In particular, if I denotes the unit matrix, then

$$|O - I| \leqq 2 \tag{1}$$

holds for every orthogonal matrix. Let a point, O, of Ω be said to lie on the "boundary," $[\Omega]$, or in the "interior," $\Omega - [\Omega]$, of Ω according as the sign of equality does or does not take place in (1); so that

$$[\Omega]: \quad |O - I| = 2; \quad \text{i.e.,} \quad \Omega - [\Omega]: \quad |O - I| < 2. \tag{2}$$

2. Let Ω_0 denote the set of those matrices which are representable in the form $e^S (= I + S + \frac{1}{2}S^2 + \cdots)$, where S is some bounded, real, skew-symmetric matrix. It is easily verified that, for every such S, the matrix e^S is orthogonal, but the converse is not true (cf. (I) below); so that Ω_0 is a proper subset of Ω. A point, O, of Ω will be called a rotation or a reflection according as it is in Ω_0 or in its complement, $\Omega - \Omega_0$. This nomenclature is suggested by the circumstance that a *finite* real, orthogonal matrix is well-known to be of positive or of negative determinant ($= \pm 1$) according as it is or is not representable as the exponential of a real, skew-symmetric matrix.

While no determinants are available for the matrices of Ω, it is clear from the above definition of Ω_0 that, if O is any orthogonal matrix,

$$ORO^{-1} \text{ is a rotation if } R \text{ is,} \tag{3}$$

* Received April 13, 1951.

and that

(4) R^{-1} is a rotation if R is.

On the other hand, it will be seen in § 11 and § 13 that, in contrast to what holds for finite matrices,

(5) R_1R_2 can be a reflection if R_1, R_2 are rotations,

and that

(6) R_1R_2 can be a reflection if R_1, R_2 are reflections.

It should be noted that (6) is not implied by (5), nor (5) by (6). On the other hand, since the orthogonal matrices form a group, (3) is equivalent to the statement that

(7) ORO^{-1} is a reflection if R is,

and (4) to the statement that

(8) R^{-1} is a reflection if R is.

It follows that either all or none of the $k!$ products of k orthogonal matrices $R_1, \cdots, R_k$ are rotations. Clearly, it is sufficient to prove this for $k = 2$, i. e., to show that if R_1R_2 is a rotation, then R_2R_1 is. But this follows by applying (3) to $R = R_1R_2$, $O = R_2$.

The following fact may also be mentioned:

(*) Neither Ω_0 nor $\Omega - \Omega_0$ is an open set (hence neither of them is a closed set) on Ω.

It is understood that these and all the subsequent topological notions refer to the topology determined by the $|O_1 - O_2|$-metric on Ω.

3. In the case of finite matrices, the set of all orthogonal matrices breaks into two closed manifolds (those of determinant $+1$ and -1, respectively), which have no point in common. In contrast, it will follow from (I) and (II) below that Ω_0 and $\Omega - \Omega_0$ contain points which are in the closures of the respective complementary sets, $\Omega - \Omega_0$ and Ω_0. Actually (I) and (II) together will supply a characterization of all these points O of "confluence" (of either kind).

4. The preceding assertions concerning the "confluence" of Ω_0 and $\Omega - \Omega_0$ neither contain nor are contained in the following fact, proved in [2]:

(9) Ω is arcwise connected.

In the case of finite orthogonal matrices, the rotations are known to form a connected manifold, and the same is true of the reflections. For the case of infinite matrices, it was shown in [2] that

(10) Ω_0 is arcwise connected,

and it will be proved in § 29 that

(11) $\Omega - \Omega_0$ is arcwise connected.

If t is a real number, then tS is a real, bounded, skew-symmetric matrix whenever S is. Hence (10) is obvious; cf. [2]. In fact, if R is a rotation, say $R = e^S$, then $R(t) = e^{tS}$, where $0 \leqq t \leqq 1$, represents a continuous path contained in Ω_0. But this path begins $(t = 0)$ at I and ends $(t = 1)$ at the given R.

5. Results corresponding to (9)-(11) will be proved for the "boundary" and for the "interior" of Ω, as defined by (2):

(12) $[\Omega]$ is arcwise connected

and

(13) $\Omega - [\Omega]$ is arcwise connected.

It will also be shown that, from the point of view of the arcwise connectivity of the boundary, $[\Omega]$, with the "center," I, of Ω, there are on $[\Omega]$ two types of points:

(A) There exist on $[\Omega]$ points O corresponding to which it is possible to find continuous paths $Q(t)$, $0 \leqq t \leqq 1$, connecting $O = Q(1)$ to $I = Q(0)$ in such a way that the given O is the only point of the path which is not in $\Omega - [\Omega]$.

(B) There exist on $[\Omega]$ points O which cannot be connected to I in the way specified under (A).

It should be noted that (A) is not implied by (13).

6. It will be easy to show that the "interior" of Ω contains none of the reflections and not all of the rotations, i. e., that

(14) $[\Omega] \supset \Omega - \Omega_0 \neq [\Omega]$.

The set of the rotations on the "boundary," i. e., the intersection, $[\Omega]\Omega_0$, of $[\Omega]$ and Ω_0, has quite an involved structure on $[\Omega]$. In fact, it turns out that, while

(15) $[\Omega]\Omega_0$ is arcwise connected,

it is not dense on $[\Omega]$ (or on Ω_0), and it is neither open nor closed on $[\Omega]$.

In case of a finite dimension number, all rotations form an invariant subgroup (of index 2) of all orthogonal transformations, and so any fixed co-set supplies a topological mapping of the space of all rotations on that of all reflections. In view of (5) or (6), this argument cannot be applied in the present case. But it will remain undecided whether Ω_0, nevertheless, is topologically equivalent to $\Omega - \Omega_0$. If it should be, then the "dual" results are mere corollaries; for instance, (11) then is equivalent to (10) and, therefore, trivial. It would, of course, be sufficient (but not necessary) to assure the existence of *some* reflection, say R_0, having the property that R_0O is a reflection or a rotation according as O is a rotation or a reflection.

The Spectral Characterization of Rotations and Reflections.

7. The proofs will depend on the spectral resolution of unitary matrices [3], pp. 268-277, applied in [1] to the real subgroup of the unitary group.

The purpose of this chapter is a spectral characterization of the rotations (hence, of the reflections as well). In view of (3) (or of (7)), it will always be allowed to assume that the orthogonal matrix to be considered is given in any of its normal forms which can be attained by orthogonal transformations.

In particular, it can be assumed that the contribution of the continuous spectrum (if any) has been split off by an orthogonal transformation. After such a transformation, every orthogonal matrix appears in exactly one of the forms.

$$(16) \qquad O = C; \qquad O = P_\infty; \qquad O = P_n \dot{+} C, \qquad 1 \leqq n \leqq \infty,$$

where C and P_∞ denote infinite orthogonal matrices having no point spectra and no continuous spectra, respectively, while if $n < \infty$, then P_n denotes a finite, n-rowed orthogonal matrix. It is understood that, whether $n < \infty$ or $n = \infty$ in the third of the cases (16), the symbol $\dot{+}$ means this: $x = 0$ is the only common point of those two linear subspaces of the x-space $\mathfrak{R}$ on which C and P_n operate.

If ϕ is any (real) angle, let $B = B(\phi)$ denote the binary matrix representing rotation by ϕ in a plane. Thus $B(0)$ is the two-rowed unit matrix, and $B(\pi)$ the negative of it. Hence, if the multiplicity with which -1 occurs in the spectrum of P_n is either finite and even (possibly 0) or infinite, then P_n is orthogonally equivalent to a matrix which, when denoted simply by P_n, is of the form

$$(17) \quad P_n = Q_n \text{ or } P_n = 1 \dot{+} Q_n, \text{ where } Q_n = B(\phi_1) \dot{+} B(\phi_2) \dot{+} \cdots, \quad (n \leqq \infty),$$

and (17) must be replaced by

$$(18)\quad P_n = R_n \text{ or } P_n = 1 \dotplus R_n, \text{ where } R_n = -1 \dotplus B(\phi_1) \dotplus B(\phi_2) \dotplus \cdots, \quad (n \leqq \infty),$$

if -1 occurs in the spectrum of P_n with a multiplicity which is finite and odd (the angles ϕ_m need not be distinct). It is understood that 1 and -1 in (17)-(18) represent the one-rowed unit matrix and its negative; that, whether $n < \infty$ or $n = \infty$, the number of the B-terms is $\frac{1}{2}n$ or $\frac{1}{2}(n-1)$ in (17) according as n is even or odd, and is $\frac{1}{2}(n-1)$ or $\frac{1}{2}(n-2)$ in (18) according as n is odd or even; finally, that the alternative cases are needed in (17) as well as in (18) in order to take care of matrices P_n, where $n \leqq \infty$, in which the multiplicity with which $+1$ occurs in the spectrum is not or is finite and odd.

8. It follows from (IV) in § 28 below that, if R is an orthogonal matrix not containing -1 in its point spectrum, and if $E(\lambda)$, where $0 \leqq \lambda \leqq 2\pi$, denotes the spectral matrix of $-R$, then the matrix

$$(19)\qquad R_t = \int_0^{2\pi} e^{it(\lambda-\pi)} dE(\lambda), \text{ where } 0 \leqq t \leqq 1,$$

is orthogonal and satisfies the functional equation $R_u R_v = R_{u+v}$. Since (19) also implies that $R_1 = -Re^{-i\pi} = R$, it follows that $R_{\frac{1}{2}} R_{\frac{1}{2}} = R$. Accordingly, if -1 does not occur in the point spectrum of an orthogonal matrix, R, then R is the square of some orthogonal matrix, $R_{\frac{1}{2}}$.

9. This fact will be combined with the following

LEMMA. *A matrix is a rotation if and only if it is the square of some orthogonal matrix.*

Since e^S is the square of $e^{\frac{1}{2}S}$, only the first of the two assertions of this Lemma needs a proof. But M. H. Martin has proved ([1], p. 590) that every orthogonal matrix, O, can be factored (not in a unique way) as follows:

$$(20)\qquad O = Te^S, \text{ where } TS = ST \text{ and } T^2 = I.$$

Here I is the unit matrix, and T, S are two real, bounded matrices the first of which is symmetric while the second is skew-symmetric. Since (20) implies that $O^2 = e^{2S}$, and since $2S$ is skew-symmetric, it follows that every O^2 is a rotation.

10. It will now be easy to prove the main theorem on rotations:

(I) *An orthogonal matrix is a rotation if and only if the multiplicity with which* -1 *occurs in its point spectrum is either finite and even (possibly* 0) *or* ∞.

First, let O be a rotation and suppose, if possible, that -1 occurs in its point spectrum with a finite, odd multiplicity. After an orthogonal transformation, it can be assumed that O is given in the form

$$O = \begin{pmatrix} -I^{2h+1} & 0 \\ 0 & R \end{pmatrix},$$

where I^{2h+1} is the $(2h+1)$-rowed unit matrix and R does not contain -1 in its point spectrum. Since O is a rotation, there exists a real, bounded skew-symmetric matrix S such that

$$O = e^S, \qquad S = \begin{pmatrix} A & B \\ -B' & C \end{pmatrix}$$

where the prime denotes the operation of transposition and A is a $(2h+1)$-rowed square matrix. It follows from $Se^S = e^S S$ that $RB' = -B'$. Hence $B' = 0$, and so

$$e^S = \begin{pmatrix} e^A & 0 \\ 0 & e^C \end{pmatrix}.$$

The last three formula lines imply that $e^A = -I^{2h+1}$. But this contains a contradiction, since $\det e^A > 0$ but $\det -I^{2h+1} < 0$. This proves the second assertion of (I).

In view of (16) and of the above Lemma, the proof of (I) will be complete if it is shown that every matrix C is the square of some orthogonal matrix, and that the same is true of every matrix (17). But every binary rotation, $B(\phi)$, is the square of such a rotation, $B(\frac{1}{2}\phi)$, while the one-rowed rotation matrix 1 is its own square. Consequently, it is sufficient to show that every C is the square of an orthogonal matrix. But C denotes an orthogonal matrix having a continuous spectrum only, hence -1 is surely not in the point spectrum of C, i. e., $R = C$ is of the type considered in § 8 before the Lemma and is therefore the square of the corresponding $R_{\frac{1}{2}}$.

11. It was shown in [2] that every reflection is the product of three rotations, and the question was raised whether two rotations would not always suffice. It will now be shown that (I) implies that the answer to this question is affirmative, and can even be refined as follows:

(†) *Every reflection can be represented as the product of two rotations; in addition, the latter can be chosen so as to be commutable.*

The assertion of this corollary, (†), of (I) is invariant under an arbitrary orthogonal transformation of a given reflection. It can therefore be assumed that if the latter is denoted by O and if k, where $0 \leqq k \leqq \infty$, denotes the multiplicity with which 1 occurs in the point spectrum of O, then $O = I^k \dot{+} R$, where I^k denotes the k-rowed unit matrix, 1 is not in the point spectrum of R, and either of the terms I^k, R may be absent. In the proof of (†), two cases have to be distinguished, according as $k < \infty$ or $k = \infty$.

If $k < \infty$, then R is infinite, and so the assertion of (†) follows by writing $O = I^k \dot{+} R$ as the product of the two matrices $I^k \dot{+} (-R)$, $I^k \dot{+} (-I^\infty)$. In fact, (I) shows that both of these matrices are rotations, since, by assumption, -1 occurs with the multiplicity 0 in the point spectrum of $I^k \dot{+} (-R)$, and with the multiplicity ∞ in the point spectrum of $I^k \dot{+} (-I^\infty)$. In the remaining case, where $k = \infty$, it is sufficient to write O as the product of $-O$ and $-I^\infty$, since -1 occurs in the point spectra of the latter two matrices with the multiplicity ∞. This proves (†).

Clearly, (†) and (10) imply (9). It is also clear that (5) follows from (†), and therefore from (I).

In contrast, (6) will depend on (II) below. It will remain undecided whether, corresponding to the refinement (†) of (5), it is possible to refine (6) to the statement that every reflection is the product of two reflections.

The Closures of the Rotations and of the Reflections.

12. For the sake of brevity, let an orthogonal matrix, O, be said to have a *pure vicinity* (*in* Ω) if there exists a positive $\beta = \beta(O)$ having the following property: Either every orthogonal matrix, Q, satisfying $|Q - O| < \beta$ is a rotation or every such Q is a reflection (according as O itself is a rotation or a reflection, $|Q - O| < \beta$ being satisfied by $Q = O$). Every orthogonal matrix is of one of two possible types, in accordance with the following theorem:

(II) *An orthogonal matrix, O, has a pure vicinity (in Ω) or is a cluster point both of rotations and reflections according as -1 is not or is in the essential spectrum of O.*

Corollary. *An orthogonal matrix, O, has a pure vicinity (in Ω) if and only if -1 is not in the essential spectrum of O.*

Here and in the sequel, the *essential spectrum* of an orthogonal matrix,[1] O, is meant to be the set of those λ-values which are either cluster points of the spectrum of O or are in the point spectrum of O with an infinite multiplicity (or both). For instance, $\lambda = 1$ is in the essential spectrum of I. Since the continuous spectrum of every O is a perfect set (unless it is vacuous), a λ-value can belong to the essential spectrum for any one, for any two, or for all three, of the following reasons: λ is in the continuous spectrum; λ is a cluster point of the point spectrum; λ is an eigenvalue of infinite multiplicity. Since O is an infinite matrix, its essential spectrum always contains at least one point. Needless to say, the essential spectrum is a subset of the spectrum (in fact, the latter is a closed set). The notions of the point spectrum, the continuous spectrum, etc. are meant, of course, in terms of the spectral theory of unitary matrices ([3], pp. 268-277).

13. Theorem (II), the proof of which will be lengthy, supplies a short proof of (6). To this end, let D denote a diagonal matrix which differs from the infinite unit matrix I only in that a single diagonal element of I is changed to -1. Thus $-D$ is a rotation containing -1 in its essential spectrum. It follows therefore from (II) that there exists a sequence of reflections $L_1, L_2, \cdots$ such that $L_m \rightarrow -D$ as $m \rightarrow \infty$ (the convergence refers to the metric of Ω). Similarly, since the negative, $-I$, of the unit matrix is, by (I), a rotation, and since -1 is in its essential spectrum, it follows from (II) that there exists a sequence of reflections $K_1, K_2, \cdots$ such that $K_m \rightarrow -I$. Consequently $L_m K_m$ tends to $-D$ times $-I$, i. e., $L_m K_m \rightarrow D$. But (I) and the definition of D show that D is a reflection not containing -1 in its essential spectrum. It follows therefore from (II) that $L_m K_m \rightarrow D$ is possible only if $L_m K_m$ is a reflection from a certain m onward, say for every $m \geqq j$. Hence, a pair of orthogonal matrices satisfying (6) follows by choosing $R_1 = L_j$, $R_2 = K_j$.

As another application of (I) and (II), the assertions of (*) at the end of § 2 will now be proved. First, it is clear from (I) that there exist both rotations and reflections for which -1 is in the essential spectrum. It follows therefore from (II) that neither the set of all rotations nor that of all reflections is an open or a closed set.

---

[1] In the Hermitian case, the idea and certain fundamental applications of essential spectrum are due to H. Weyl, *Rend. Palermo*, vol. 27 (1909), pp. 373-392, and *Mathematische Annalen*, vol. 68 (1910), p. 251.

14. There will now be collected the tools needed from spectral theory.

If H is any bounded, Hermitian matrix, then e^{iH} is unitary. Conversely, it was shown in [3], pp. 268-277, that every unitary matrix is representable in the form e^{iH}, even if the spectrum of H is restricted to the interval $0 \leqq \lambda \leqq 2\pi$, and even if $\lambda = 2\pi$ is restricted to be not in the point spectrum of H; furthermore, there belongs to every unitary matrix essentially one H subject to these restrictions. Thus if $E(\lambda)$ denotes the spectral matrix of H, then

$$(21a) \qquad E(\lambda - 0) = E(\lambda), \; -\infty < \lambda < \infty,$$

and

$$(21b) \qquad E(\lambda) = \| 0 \| \text{ if } \lambda \leqq 0, \text{ and } E(\lambda) = I \text{ if } \lambda \geqq 2\pi.$$

In his thesis, M. H. Martin has specified those $E(\lambda)$'s which satisfy (21a)-(21b) and have the property that, if H is the Hermitian matrix having the spectral matrix $E(\lambda)$, then the unitary matrix e^{iH} is real, i. e., orthogonal. The result of his reality discussion ([1], pp. 598-600) is as follows: e^{iH} is orthogonal if and only if

$$(22a) \qquad \bar{E}(+0) = E(+0), \qquad (E(0) = \| 0 \|),$$

and there exists a real constant a having the property that

$$(22b) \qquad \bar{E}(\lambda) + E(2\pi - \lambda) = aI \text{ for } 0 < \lambda < 2\pi$$

holds at those interior points λ at which $\bar{E}(\lambda) + E(2\pi - \lambda)$ is continuous. The bars in (22a)-(22b) denote complex conjugation, i. e., ordinary transposition, of the Hermitian matrix E.

15. This machinery will not be needed in the proof of the *first* assertion of (II), which proceeds as follows:

Suppose that O is an orthogonal matrix not containing -1 in its essential spectrum, and let k, where $0 \leqq k < \infty$, be the multiplicity with which -1 is in the point spectrum ($k = \infty$ is impossible, since -1 is not in the essential spectrum). It will be shown that if Q is any orthogonal matrix for which $| Q - O |$ is smaller than a certain positive number $\beta = \beta(O)$, and if $l = l(Q)$ denotes the multiplicity with which -1 occurs in the point spectrum of Q, then $l < \infty$ and $l \equiv k \pmod 2$. In view of (I), this will prove the first assertion of (II) (and even more, since l is claimed to be finite).

Let $x_1, \cdots, x_k$ denote a set of linearly independent eigenvectors of O belonging to -1, so that $Ox_h = -x_h$ (this x_h-set is vacuous if $k = 0$).

Since -1 is not in the essential spectrum of O, it is clear that $|(O+I)x|$ has a positive lower bound when x varies over those unit vectors which are orthogonal to the space spanned by these eigenvectors. Let such a positive lower bound be denoted by 2β. It will be shown that the $\beta = \beta(O)$ thus defined has the desired property.

In order to see this, let Q be any orthogonal matrix satisfying $|Q-O| < \beta$. Then -1 is not in the essential spectrum of Q and *at most* k eigenvalues[2] of Q are in the λ-circle $|\lambda+1| \leqq \beta$. For suppose the contrary. Then there exists a unit vector, x, which satisfies the inequality $|(Q+I)x| \leqq \beta$ and is orthogonal to the space spanned by the k eigenvectors, $x_1, \cdots, x_k$, of O considered above. But $|(O-Q)x| < \beta$, since $|x| = 1$ and $|O-Q| < \beta$. It follows therefore from $|(Q+I)x| \leqq \beta$ that $|(O+I)x| < 2\beta$. This contradicts, however, the definition of 2β as a lower bound.

If the rôles of O, Q are interchanged in the preceding proof, what follows is that Q has *at least* k eigenvalues in the λ-circle $|\lambda+1| \leqq \beta$. Consequently, Q has *exactly* k eigenvalues (and no essential spectrum) in this circle. But if λ is a complex eigenvalue of Q, then, since Q is orthogonal (hence real), the complex conjugate of λ is an eigenvalue of the same multiplicity as λ. Since the circle $|\lambda+1| \leqq \beta$ is bisected by the real axis and since the spectrum of Q is on the circle $|\lambda| = 1$, it follows that $\lambda = -1$ occurs in the point spectrum of Q with a multiplicity, say $l(\geqq 0)$, for which the difference $k-l$ must become even. As explained above, this proves the first of the two assertions of (II).

16. In order to prove the remaining assertion of (II), suppose that -1 is in the essential spectrum of an orthogonal matrix O. The assertion to be proved is that both rotations and reflections must then cluster at the point O of Ω.

Since this assertion remains unaltered if O is replaced by a matrix orthogonally equivalent to O, it can be assumed that O is in one of its three normal forms (16), with (17)-(18). But -1 then is in the essential spectrum of O either because it is in the spectrum of C or because -1 is either a cluster point or a point of infinite multiplicity in the point spectrum of P (the latter P can belong either to the second or to the third of the three cases in (16)).

[2] By eigenvalues are meant points in the point spectrum and, when estimating their number, they are meant to be enumerated so as to take into account their multiplicities.

An orthogonal matrix will be said to be of type (*) if -1 occurs in its point spectrum with a finite (possibly zero) multiplicity. Let O_1 and O_2 be of type (*); then so also is $O_1 \dot{+} O_2$, and (I) implies the following fact: If O_1 is a rotation (reflection), then the orthogonal matrix $O_1 \dot{+} O_2$ is a rotation (reflection) if and only if O_2 is a rotation. Hence it is easy to see from (16) and (17)-(18) that it is sufficient to show that O is a cluster point both of rotations and of reflections, *of type* (*), in the following three particular cases:

(i) O has no point spectrum and -1 is in the (continuous) spectrum.

(ii) O has no continuous spectrum and -1 is a cluster point of the (point) spectrum but is not in the point spectrum.

(iii) $O = -(1 \dot{+} 1 \dot{+} \cdots)$, i. e., $O = -I$.

(Moreover, it is seen that a portion of the assertion of (II) can be improved as follows: If -1 is in the essential spectrum of an orthogonal matrix O, then O is a cluster point both of rotations and reflections of type (*).)

Case (iii) is straightforward. In fact, if $0 < \epsilon < \pi$, and if $B(\phi)$ denotes again the binary rotation by the angle ϕ, then it is seen from (I) that

$$P_\epsilon = B(\pi - \epsilon/1) \dot{+} \beta(\pi - \epsilon/2) \dot{+} B(\pi - \epsilon/3) \dot{+} \cdots$$

is a rotation, and that $-1 \dot{+} P_\epsilon$ is a reflection. On the other hand, since $1/1^2 + 1/2^2 + \cdots < \infty$, and since $-B(\pi)$ is the binary unit matrix, it is clear that both the rotation P_ϵ and the reflection $-1 \dot{+} P_\epsilon$ are of type (*) and tend, as $\epsilon \to 0$, to $-I$, which is the O of case (iii).

Case (ii) can be disposed of similarly. First, it is seen from (17) that a normal form of O in the case (ii) is

(24) $\quad O = B(\phi_1) \dot{+} B(\phi_2) \dot{+} \cdots$ or $O = 1 \dot{+} B(\phi_1) \dot{+} B(\phi_2) \dot{+} \cdots,$

where the angles $\phi_1, \phi_2, \cdots$ satisfy the following pair of conditions: $|\phi_m| < \pi$ holds for every m, and $\liminf |\pi - \phi_m| = 0$ as $m \to \infty$. In view of the *first* of these conditions, (I) and (24) show that O is a rotation, and that it remains a rotation (and of type (*)) if one term of (24), say the first, is changed from $B(\phi_1)$ to $B(\phi_1 \pm \epsilon)$, provided that $0 < \epsilon < \pi - |\phi_1|$. But if O_ϵ denotes the matrix which thus results from (24), then $O_\epsilon \to O$ as $\epsilon \to 0$, and so O is a cluster point of rotations of type (*). In order to prove that O is a cluster point of reflections of type (*) as well, use must be made of the *second* of the conditions, which is $\liminf |\pi - \phi_m| = 0$ as $m \to \infty$.

Since the latter condition implies that $\phi'_m \to \pi$ holds for a suitable

infinite subsequence $\phi'_1, \phi'_2, \cdots$, of $\phi_1, \phi_2, \cdots$, it is clear that, with reference to every $\epsilon > 0$, it is possible to write (24) in the form

$$(25) \qquad O = O_\epsilon \dotplus (-I + A_\epsilon),$$

where the plus sign occurring in $-I + A_\epsilon$ refers to *ordinary* matrix addition, $-I$ is $B(\pi) \dotplus B(\pi) \dotplus \cdots$, and the infinite matrix A_ϵ represents a correction term, with $|A_\epsilon| < \epsilon$. It follows from (I), and from the fact that O is in case (ii), that O_ϵ is a rotation not containing -1 in its point spectrum. It follows therefore from the remarks preceding the statements of cases (i)-(iii) that O will be proved to be a cluster point of reflections of type (*) if it is ascertained that, for every $\epsilon > 0$, the inequality $|(-I + A_\epsilon) - R| < 2\epsilon$ can be satisfied by a certain reflection $R = R(\epsilon)$. But since $|A_\epsilon| < \epsilon$, the existence of such an $R = R(\epsilon)$ follows if $|-I - R_m| \to 0$, as $m \to \infty$, holds for some sequence of reflections $R_1, R_2, \cdots$ of type (*). Since the existence of such reflections R_m was proved in the treatment of case (iii), the treatment of case (ii) is now complete.

17. The treatment of the remaining case, (i), will now be reduced to the following

LEMMA. *If -1 is in the spectrum, but not in the point spectrum, of an orthogonal matrix O, then there exists a sequence of orthogonal matrices $O_1, O_2, \cdots$ satisfying $|O - O_m| \to 0$, as $m \to \infty$, and having the property that -1 is a cluster point of the point spectrum, but is not in the point spectrum, of O_m, where $m = 1, 2, \cdots$.*

Needless to say, the matrices O and $O_1, O_2, \cdots$ are rotations by necessity: cf. (I).

It is clear that if an O is in case (i), then it satisfies the assumptions of the Lemma (but the converse is not true). If a matrix O_m of the above sequence does not possess any continuous spectrum, then it is in case (ii), so that both rotations and reflections (of type (*)), say R_k, must cluster at O_m. The last assertion remains true even if O_m does possess a continuous spectrum. In fact, if $O_m = C \dotplus P_\infty$ (in accordance with (16); note that -1 is a cluster point of the point spectrum), then P_∞ is in case (ii), so that both rotations and reflections R_k of type (*) cluster at P_∞; hence both rotations and reflections $C \dotplus R_k$ of this same type cluster at O_m; cf. the remarks preceding the statements of (i)-(iii) in § 16. It follows therefore from the Lemma that every O of the case (i) must be a cluster point (on Ω) of rotations and reflections of type (*). Accordingly, the proof of (II) will be complete if the Lemma is proved.

18. To this end, let O be a matrix satisfying the assumptions of the Lemma. Then 1 is in the essential spectrum, but not in the point spectrum, of the orthogonal matrix $-O$. Hence, if $E(\lambda)$ denotes the spectral matrix, satisfying (21a)-(22b), of O, then $E(\lambda)$ is not constant on any interval $0 < \lambda < \eta$ and is continuous at $\lambda = 0$. In view of the parenthetical remark in (22a), the latter property means that

$$E(+0) = E(0) = \| 0 \|. \tag{26}$$

On the other hand, the non-constancy of $E(\lambda)$ on any interval $0 < \lambda < \eta$ means that, with reference to every positive ϵ which is less than π, it is possible to choose a sequence of values $\lambda = \lambda_n = \lambda_n(\epsilon)$ satisfying

$$\epsilon = \lambda_0 > \lambda_1 > \cdots > \lambda_n \to 0, \text{ where } n \to \infty, \tag{27}$$

and

$$E(\lambda_n) \neq E(\lambda_{n+1}), \text{ where } n = 0, 1, 2, \cdots. \tag{28}$$

Disregarding an enumerable set of λ-values, one may assume that $\lambda_0, \lambda_1, \lambda_2, \cdots$ are continuity points of $E(\lambda) + E(2\pi - \lambda)$.

With reference to the sequence (27), where $\lambda_n = \lambda_n(\epsilon)$, define for $-\infty < \lambda < \infty$ a matrix function $F(\lambda) = F_\epsilon(\lambda)$ as follows:

$$F(\lambda) = E(\lambda_n) \text{ if } \lambda_{n+1} < \lambda \leqq \lambda_n \text{ and } F(2\pi - \lambda) = E(2\pi - \lambda_n) \quad \text{if } \lambda_{n+1} \leqq \lambda < \lambda_n \tag{29}$$

and

$$F(\lambda) = E(\lambda) \text{ if } \epsilon < \lambda \leqq 2\pi - \epsilon \qquad (\epsilon = \lambda_0), \tag{30}$$

finally $F(\lambda) = \| 0 \|$ if $\lambda \leqq 0$ and $F(\lambda) = I$ if $\lambda \geqq 2\pi$. Then it is seen from (26) and (21a)-(22b), where $E(\lambda)$ is the spectral matrix of a Hermitian matrix, that

$$F(+0) = F(0) = \| 0 \|, \qquad F(2\pi - 0) = F(2\pi) = I \tag{31}$$

and that $F(\lambda)$ is the spectral matrix of the Hermitian matrix

$$H = \int_{-\infty}^{\infty} \lambda dF(\lambda) = \int_{+0}^{2\pi - 0} \lambda dF(\lambda), \tag{32}$$

and therefore that of the unitary matrix e^{iH}.

Clearly, $F(\lambda)$ is a step-function on the outside of the interval $\epsilon < \lambda < 2\pi - \epsilon$, with jumps possible, for $\lambda \leqq \epsilon$, only at the λ-values (27), and which, in view of (28), take place at each of the values (27), clustering at 0. Hence e^{iH} has no continuous spectrum near the value $e^{i0} = 1$, and its

point spectrum clusters at the value $e^{i0} = 1$, which value is not in the point spectrum of e^{iH}.

In addition, since conditions (21a)-(22b) are satisfied by the given $E(\lambda)$, it is seen from (29), (30) and (31) and the properties of the sequence $\lambda_0, \lambda_1, \lambda_2, \cdots$ that these conditions are satisfied if E is replaced by F. But the satisfaction of (22a)-(22b) by $E = F$ means that the unitary matrix having the spectral matrix $F(\lambda)$ is real, i. e., orthogonal.

Accordingly, e^{iH} is an orthogonal matrix which has no continuous spectrum near the value 1, and which contains 1 is in its essential spectrum but not in its point spectrum.

19. By construction, H is here a function, H_ϵ, of ϵ, where $0 < \epsilon < \pi$. It will now be shown that

$$(33) \qquad | e^{iH} + O | \to 0 \text{ as } \epsilon \to 0 \text{ in } H = H_\epsilon.$$

First, since $E(\lambda)$ and $F(\lambda)$ are spectral matrices,

$$(34) \qquad \left| \int_0^\epsilon \lambda dE(\lambda) - \int_0^\epsilon \lambda dF(\lambda) \right| \leqq \epsilon + \epsilon.$$

In virtue of (26), (31) and (21a) (the last being valid for E itself and for $E = F$),

$$\exp \int_0^\epsilon i\lambda dE(\lambda) = I - E(\epsilon) + \int_0^\epsilon e^{i\lambda} dE(\lambda)$$

holds for E itself and for $E = F$. It is seen from (27). (29) and (34) that, as $\epsilon \to 0$,

$$\int_0^\epsilon e^{i\lambda} dE(\lambda) = \int_0^\epsilon e^{i\lambda} dF(\lambda) = o(1).$$

Disregarding an enumerable set of ϵ-values, choose ϵ so that $\lambda = \epsilon$ is a continuity point of $E(\lambda)$ and of $E(2\pi - \lambda)$ (and hence of $F(\lambda)$ and of $F(2\pi - \lambda)$; cf. (29) and (30)). Then, by the last formula line and (22b), valid for $E = F$ as well as for E itself, the relation

$$\int_{2\pi-\epsilon}^{2\pi} e^{i\lambda} dE(\lambda) - \int_{2\pi-\epsilon}^{2\pi} e^{i\lambda} dF(\lambda) = o(1), \qquad \epsilon \to 0,$$

holds; cf. (22b), valid for E itself and for $E = F$, and note that the present

$E(\lambda)$ and $F(\lambda)$ are continuous at $\lambda = 0$ as well as at $\lambda = 2\pi$. Finally, from (30),

$$\int_{\epsilon+0}^{2\pi-\epsilon-0} e^{i\lambda} dE(\lambda) - \int_{\epsilon+0}^{2\pi-\epsilon-0} e^{i\lambda} dF(\lambda) = 0.$$

Addition of the last three formula lines shows that

$$(35) \qquad \int_{0}^{2\pi} e^{i\lambda} dE(\lambda) - \int_{0}^{2\pi} e^{i\lambda} dF(\lambda) = o(1)$$

as $\epsilon \to 0$, where $F(\lambda) = F_\epsilon(\lambda)$. But the difference on the left of (35) is identical with $-O - e^{iH}$, since $E(\lambda)$ and $F(\lambda)$ have been defined as the spectral matrices of $-O$ and e^{iH}, respectively. Consequently, (35) proves (33).

20. If $m = 1, 2, \cdots$, let O_m denote the matrix $-e^{iH}$ which results if ϵ in $H = H_\epsilon$ is chosen to be $\epsilon_m (> 0)$, where $\epsilon_m \to 0$ as $m \to \infty$. Then $| O_m - O | \to 0$, by (33). On the other hand, the last sentence of § 18 shows that every O_m has all the properties claimed in the Lemma of § 17. Thus the proof of the Lemma (and hence that of (II)) is complete.

The "Boundary" $[\Omega]$ of the Ω-Space.

21. If $E(\lambda)$, where $0 \leqq \lambda \leqq 2\pi$, is the spectral matrix of an orthogonal matrix O, define the number $\gamma = \gamma(O)$, where $0 \leqq \gamma \leqq \pi$, as follows: According as $\lambda = \pi$ is not or is a point of constancy of $E(\lambda)$ (i. e., according as -1 is or is not in the spectrum of O), let γ denote 0 or the largest number ($\leqq \pi$) having the property that $E(\lambda) = E(\pi)$ whenever $| \lambda - \pi | < \gamma$. Then

$$(36) \qquad | I - O | = 2 \sin \tfrac{1}{2}(\pi - \gamma)$$

holds in either case. In fact, since

$$I - O = \int_{0}^{2\pi} (1 - e^{i\lambda}) dE(\lambda),$$

it is readily seen from the definition of $\gamma = \gamma(O)$ that $| 1 - e^{i(\pi-\gamma)} |$ is the least upper bound of the length of the vector $(I - O)x$ when x varies over all vectors of unit length. But this evaluation of l. u. b. $|(I - O)x|/|x|$ is equivalent to (36).

It is seen from the definition, (2), of the "boundary," $[\Omega]$, of Ω that the particular case $\gamma = 0$ of (36) can be formulated as follows:

(i) *An orthogonal matrix, O, represents a point on the "boundary," $[\Omega]$, if and only if -1 is in the spectrum of O.*

This criterion, (i), leads to the following corollary:

(ii) *The "boundary," $[\Omega]$, contains all reflections and some rotations.*

In fact, (I) implies that -1 is in the spectrum of O if, but not only if, O is a reflection. Hence (ii) follows from (i). Clearly, (14) is just a restatement of (ii).

22. Let F denote the diagonal matrix $-1 \dotplus D$, where D is the orthogonal matrix defined at the beginning of § 13. Then the essential spectrum of either of the matrices D, F, being the single value 1, does not contain the value -1. Furthermore, (I) implies that F is a rotation and that D is a reflection. It follows therefore from (II) that the rotation F is not in the closure of all reflections of Ω and that the reflection D is not in the closure of all rotations of Ω. But (i) and the definitions of D and F show that both D and F are in $[\Omega]$. Since $[\Omega]$ is a subset of Ω, it follows that $[\Omega]$ contains a reflection $(= D)$ which is not in the closure of the rotations contained in $[\Omega]$. It is also seen that $[\Omega]$ contains a rotation $(= F)$ which is not in the closure of the reflections contained in $[\Omega]$, since, according to (ii), all reflections are in $[\Omega]$. Accordingly, the situation is as follows:

(iii) *Neither the set of all reflections nor the set of all rotations contained in $[\Omega]$ is dense on $[\Omega]$.*

It was also seen that $[\Omega]$ contains a point $(= D)$ which fails to be in the closure of *all* rotations of Ω. Since $\Omega - [\Omega]$ is a subset of Ω, it follows that the names "boundary of Ω" and "interior of Ω" for the two sets (2) are not justifiable in terms of the topology of Ω:

(iv) *The "boundary," $[\Omega]$, of Ω is not a subset of the closure of the "interior," $\Omega - [\Omega]$, of Ω (i. e., Ω is not the closure of $\Omega - [\Omega]$).*

What is true (but trivial) is the following fact:

(v) *The "boundary," $[\Omega]$, is a closed set (i. e., the "interior," $\Omega - [\Omega]$, is an open set) on Ω.*

For, if O is any point of Ω for which $| I - O |$ is less than 2, say $2 - \epsilon$, then, since Ω is a metric space, $| I - Q | < 2 - \frac{1}{2}\epsilon (< 2)$ holds whenever

$| Q - O | < \frac{1}{2}\epsilon$. In view of (2), this means that $\Omega - [\Omega]$ is an open set. Hence (v) follows by complementation.

23. By a dualization of the content of (iii), the following facts will now be proved:

(vi) *Neither the set of the reflections nor the set of the rotations contained in* $[\Omega]$ *is an open set* (*hence, neither of these sets is a closed set*) *on* Ω.

In view of (ii), those assertions of (vi) which concern the reflections are contained in the corresponding assertions of (*), § 2, proved in § 13. Hence, in order to prove (vi), it is sufficient to show that the product $[\Omega]\Omega_0$, representing that portion of the set, Ω_0, of all rotations which is on the "boundary" of Ω, is neither an open set nor a closed set (on Ω). In other words, it is sufficient to show that the set $[\Omega]\Omega_0$ contains a point, say Q, and that the set $\Omega - \Omega_0$ (which is identical with the product of $[\Omega]$ and $\Omega - \Omega_0$) contains another point, say R, such that Q is in the closure of $\Omega - \Omega_0$ and R is in the closure of $[\Omega]\Omega_0$. The existence of Q is assured by (II) if Q is chosen to be any rotation with -1 in its essential spectrum.

In order to guarantee the existence of R one can proceed as follows: Let A denote a rotation for which -1 is in the essential spectrum but is not the point spectrum. Then $A_1 = -1 \dotplus A$ is a reflection by (I), while (II) shows there exists a sequence of reflections, B_n, such that $B_n \to A$, hence $-1 \dotplus B_n \to A_1$. Another application of (I) shows that the matrices $-1 \dotplus B_n$ are rotations (which clearly belong to $[\Omega]$). The point R may now be chosen to be A_1 and the proof of (vi) is complete.

24. In view of the parenthetical restatement in (iv), there arises the question concerning a characterization of those orthogonal matrices which are in the closure of $\Omega - [\Omega]$. The situation will be described by

(III) *A point, O, of Ω is not in the closure of $\Omega - [\Omega]$ if and only if O is in the interior of $\Omega - \Omega_0$.*

In contrast to (iv), the definition of an interior point is now based on the topology of Ω, as determined by the "strong" $| O_1 - O_2 |$-metric of § 1. In this terminology, the Corollary of (II) can of course be restated as follows:

(II bis) *Whether a point, O, of Ω be in Ω_0 or in $\Omega - \Omega_0$, it is an interior point of Ω_0 or of $\Omega - \Omega_0$ if and only if -1 is not in the essential spectrum of the orthogonal matrix represented by the point O.*

This formulation, (II bis), of (II) is adjusted to the following proof of (III).

25. Consider first that assertion of (III) in which a given O is supposed to be a point in the interior of $\Omega - \Omega_0$. Thus, in a sufficiently small vicinity of O (with reference to the "strong" $|O_1 - O_2|$-metric), there exist no rotations. Since the set $\Omega - [\Omega]$ consists only of rotations, the proof of the first assertion of (III) is complete.

The remaining assertion of (III) is that if a point, O, of Ω is not in the interior of $\Omega - \Omega_0$, then

$$(37) \qquad O \text{ is in the closure of } \Omega - [\Omega].$$

But an O is not in the interior of $\Omega - \Omega_0$ either because (a) the point O is in Ω_0 or because (b) the point O is in $\Omega - \Omega_0$ and in the closure of Ω_0. It is however clear that if (37) is granted to be true in case (a), then (37) must be true in case (b) also. Hence it is sufficient to prove (37) in case (a), i. e., under the assumption that O is a rotation. Then (I) shows that, if O is assumed to be given in its appropriate normal form (16)-(18), the P_n occurring in the last of the three possibilities (16) cannot be of type (18) and must therefore be of type (17). On the other hand, it is readily seen that (37) will be proved for the third of the three cases (16) if it is proved for the second and for the first.[3] Consequently, if the trivial term 1 mentioned in (17) is omitted, it is sufficient to prove (37) under the assumption that the rotation O is either of the form

$$(38) \qquad O = B(\phi_1) \dotplus B(\phi_2) \dotplus \cdots, \text{ where } |\phi_m - \pi| \leqq \pi,$$

or of the form $O = C$, where $B(\phi)$ is a binary rotation and C has no point spectrum.

Consider first the case $O \neq C$. Then it is clear from (38) that, corresponding to every positive ϵ which is less than π, it is possible to choose a sequence of ϕ-values $\phi_1^\epsilon, \phi_2^\epsilon, \cdots$ satisfying the following two conditions: $\epsilon \leqq |\phi_m^\epsilon - \pi| \leqq \pi$ holds for every m and $|O - O^\epsilon| \to 0$ holds as $\epsilon \to 0$, where O^ϵ denotes the orthogonal matrix $B(\phi_1^\epsilon) + B(\phi_2^\epsilon) \dotplus \cdots$. Because of the second of these conditions, (37) will follow if it is ascertained that the point O^ϵ of Ω is in $\Omega - [\Omega]$, i. e., that O^ϵ is not in $[\Omega]$. But the first of the two conditions shows that every value occurring in the spectrum of O^ϵ differs from $-1 = e^{i\pi}$ by not less than ϵ (in angular distance). Hence, -1 is not in the spectrum of O^ϵ. It follows therefore from (i), § 21, that O^ϵ is not in $[\Omega]$. This proves (37) for the case (38).

[3] This is clear when n is ∞ in the third case of (16); on the other hand, if n is finite (and hence, by (I), even), the proof will be clear from the consideration below, the treatment for P_n being essentially similar to that for P_∞.

26. Similarly, (37) will be proved for the remaining case, where $O = C$, if it is shown that there belongs to every $\epsilon > 0$ an orthogonal matrix, say O_ϵ, having the following properties: The value -1 is not in the spectrum of O_ϵ and $| C - O_\epsilon | \to 0$ holds as $\epsilon \to 0$. But the construction of such an O_ϵ hardly differs from the construction of O^ϵ in the case of (38).

First, if $E(\lambda)$, where $0 \leqq \lambda \leqq 2\pi$, denotes the spectral matrix of $-C$, then, since C has no point spectrum, $E(\lambda)$ is continuous throughout. If $0 < \epsilon < \pi$, define $F(\lambda)$ by $F(\lambda) = \| 0 \|$, $E(\lambda)$ or I according as $0 \leqq \lambda \leqq \epsilon$, $\epsilon < \lambda \leqq 2\pi - \epsilon$ or $2\pi - \epsilon < \lambda \leqq 2\pi$. Thus the conditions (21a)-(22b), which are satisfied by the given E, become satisfied by F. It is seen that (34) is also valid and the procedure is now similar to that of § 19. It is clear that $-e^{iH}$ does not have -1 in its spectrum and that $| C + e^{iH} | \to 0$ as $\epsilon \to 0$. Thus O_ϵ can be defined to be $-e^{iH}$ and the proof of (III) is complete.

Arcwise Connections.

27. If Λ is a subset of Ω and if P, Q is a pair of points in Λ, let the symbol $P(\Lambda)Q$ denote that either $P = Q$ or, if $P \neq Q$, there exists in Ω a continuous path $O = O(t)$, $a \leqq t \leqq b$ $(\geqq a)$, having the following properties: $O(t)$ is a point of Λ (for every t) and $O(a) = P$, $O(b) = Q$. Thus $Q(\Lambda)P$ implies that $P(\Lambda)Q$ and that $Q(\Lambda)R$ whenever $P(\Lambda)R$.

If $P(\Lambda)Q$ holds for every pair of points P, Q contained in Λ, then Λ is called arcwise connected. This is the meaning of the six assertions (9)-(13), (15) of the introduction, which have not been used thus far. The purpose of this chapter is to prove the last four of those six assertions. The first two of them have already been proved (cf. § 4).

The proofs will involve the circumstance that if Λ is any of the four sets in question, then Λ is an invariant subset of Ω (simply because the same is true of both $\Lambda = \Omega_0$ and $\Lambda = [\Omega]$). In other words, all the subsets Λ of Ω which are to be considered have the property

$$RAR^{-1} = \Lambda, \tag{39}$$

where R is *any* orthogonal matrix (so that R need not be in Λ). It is clear from (9) that $R(\Omega)I$. hence $O(\Lambda)ROR^{-1}$, so that, in order to prove that an O-set Λ satisfying (39) is arcwise connected, it is sufficient to exhibit in Λ a single point O_0 having the property that $ROR^{-1}(\Lambda)O_0$ holds for every point O of Λ and some point R of Ω (where R, in contrast to the fixed O_0, is a function of O). In particular, it is allowed to assume that O is given in an orthogonal normal form, (16)-(18), even though such a form of a given

matrix of Λ cannot in general be attained by using orthogonal transformations contained in Λ.

28. The proofs will be deduced from the following theorem:

(IV) *Let R be an orthogonal matrix, $E(\lambda)$ the spectral matrix of $-R$, finally t a parameter on the range $0 \leqq t \leqq 1$. Then*

(α) *if -1 is not in the point spectrum of R, the matrix defined by*

$$(40) \qquad R_t = \int_0^{2\pi} e^{it(\lambda-\pi)} dE(\lambda); \qquad 0 \leqq t \leqq 1, \qquad (R_1 = R_0, R_0 = I),$$

is orthogonal, and -1 is not in its point spectrum (for any t);

(β) *if -1 is not in the spectrum of R, it is not in the spectrum of R_t;*

(γ) *under the assumption made in (α),*

$$(41) \qquad R_u R_v = R_{u+v}.$$

Since the assumption in (α) means (and therefore the assumption in (β) implies) that 1 is not in the point spectrum of $-R$, both $\lambda = 0$ and $\lambda = 2\pi$ are continuity points of the spectral matrix, $E(\lambda)$, of $-R$, and so

$$(42) \qquad \int_0^{2\pi} = \int_{+0}^{2\pi-0}$$

in (40). (Actually, $\lambda = 2\pi$ is always a continuity point for any spectral matrix $E(\lambda)$ under the normalization (21a).) It is readily verified from (21a)-(22b) that the matrix (40), where $0 \leqq t \leqq 1$, is real. But (40) surely is a unitary matrix (in fact, $E(\lambda)$ is a spectral matrix for $0 \leqq \lambda \leqq 2\pi$). Consequently, (40) is an orthogonal matrix, as claimed by the first part of (α).

In order to prove the second claim of (α), suppose, if possible, that -1 is in the point spectrum of R_t (for some t). This means that (for that t) there exists a unit vector x satisfying $(I + R_t)x = 0$. But it is seen from (40) and (42) that

$$(43) \qquad |(I + R_t)x|^2 = \int_{+0}^{2\pi-0} |1 + e^{it(\lambda-\pi)}|^2 \, d\,|E(\lambda)x|^2$$

holds for every unit vector x, not only for those satisfying the assumption $(I + R_t)x = 0$. Hence, the latter implies the vanishing of the integral (43). But

$$(44) \qquad \int_{+0}^{2\pi-0} d\,|\,E(\lambda)x\,|^2 = \int_0^{2\pi} d\,|\,E(\lambda)x\,|^2 = |\,x\,|^2$$

is an identity in x. On the other hand, it is seen from the assumption $0 \leqq t \leqq 1$ that the function integrated in (43) is positive at every point λ of the (open) interval of integration. Consequently, the vanishing of the integral (43) implies that $x = 0$. Since this contradicts the assumption $|\,x\,| = 1$, the proof of (α) is complete.

In order to prove (β), suppose that -1 is not in the spectrum of R. This means that $\lambda = 0$ and $\lambda = 2\pi$ are points of constancy (rather than, as in the more general case of (α), just points of continuity) of $E(\lambda)$, i. e., that (42) can be improved to

$$(45) \qquad \int_0^{2\pi} = \int_\mu^{2\pi-\mu}, \text{ where } 0 < \mu < \pi.$$

But $0 < \mu < \pi$ and $0 \leqq t \leqq 1$ imply that the function integrated in (43) has a positive minimum on the interval $\mu \leqq \lambda \leqq 2\pi - \mu$. It follows therefore from (43), (44) and (45) that $|\,(I + R_t)x\,|^2 \geqq$ const. $|\,x\,|^2$, where the const. is positive and independent of x. In other words, the quadratic form $|\,(I + R_t)x\,|^2$ is positive definite. But a classical criterion of Toeplitz (cf., e. g., [4], p. 138) implies that if A is a real, bounded, normal matrix, then A^{-1} exists (as a unique, bounded reciprocal) if and only if the quadratic form $|\,Ax\,|^2$ is positive definite. Consequently, $I + R_t$ has a bounded reciprocal matrix. This proves part (β) of (IV).

Finally, in order to prove (γ), it is sufficient to observe that, since $E(\lambda)$ is a spectral matrix on $0 \leqq \lambda \leqq 2\pi$, the (unitary) matrix

$$W_t = \int_0^{2\pi} e^{i\lambda t} dE(\lambda) \quad \text{satisfies} \quad W_{u+v} = \int_0^{2\pi} e^{i\lambda u} e^{i\lambda v} dE(\lambda) = W_u W_v,$$

and that $R_t = e^{-i\pi t} W_t$, by (40).

29. *Proof of* (11). Let I^{2h-1}, where $h = 1, 2, \cdots$, denote the $(2h-1)$-rowed unit matrix. Then (I) shows that an orthogonal matrix is a reflection if and only if it is orthogonally equivalent to a matrix of the form $-I^{2h-1} \dotplus R$, where R denotes a rotation not containing -1 in its point spectrum. Since $-1 \dotplus I = -I^1 \dotplus I$ is a reflection, it follows from the remarks made at the end of § 27 that (11) is equivalent to the following statement:

$$(46) \qquad (-I^{2h-1} \dotplus R)\,(\Lambda)\,(-I^1 \dotplus I), \text{ where } \Lambda = \Omega - \Omega_0.$$

Here R denotes any rotation not containing -1 in its point spectrum. But part (a) of (IV) shows that any such R can be connected to I by a continuous path in such a way that -1 is not contained in the point spectrum of any matrix, R_t, representing a point of the path. Since this means that every point of the path R_t is a rotation, every point of the path $-I^{2h-1} \dot{+} R_t$ is a reflection, i. e., a point of $\Omega - \Omega_0$. But the end points, $t = 0$ and $t = 1$, of the latter path are $-I^{2h-1} \dot{+} I$ and $-I^{2h-1} \dot{+} R$. Consequently,

$$(-I^{2h-1} \dot{+} R)\,(\Omega - \Omega_0)\,(-I^{2h-1} \dot{+} I).$$

Hence, in order to prove (46), it is sufficient to ascertain the truth of the following relation:

$$(47) \qquad (-I^{2h-1} \dot{+} I)\,(\Lambda)\,(-I^1 \dot{+} I), \text{ where } \Lambda = \Omega - \Omega_0.$$

In order to prove (47), consider the matrix

$$(48) \qquad -I^1 \dot{+} B(\pi t) \dot{+} \cdots \dot{+} B(\pi t) \dot{+} I,$$

in which the number of B-terms is chosen to be $h - 1$ ($< \infty$, possibly 0). Since $B(\phi)$ denotes the binary matrix representing a rotation by the angle ϕ, it is clear that (48) is a reflection (i. e., a point of $\Omega - \Omega_0$) for every t, and that (48) becomes $-I^1 \dot{+} I$ at $t = 0$ and $-I^{2h-1} \dot{+} I$ at $t = 1$. This proves (47). Hence the proof of (11) is now complete.

30. *Proof of* (12). Let (15), to be proved below, be granted. Then it is clear from (11) and (14) that, in order to prove (12), it is sufficient to exhibit on $[\Omega]$ *one* rotation and *one* reflection, say O_1 and O_2, which can be joined by a continuous path contained in $[\Omega]$. But (i) in § 21 shows that both matrices $-1 \dot{+} (\pm I)$ are in $[\Omega]$ and, according to (I), only one of them is a reflection. Consequently, (12) will follow if it is ascertained that

$$(49) \qquad O^-([\Omega])O^+, \text{ where } O^{\pm} = -1 \dot{+} (\pm I).$$

But it is seen from (i) in § 21 that the truth of (49) is equivalent to the truth of

$$(50) \qquad -I(\Omega)I.$$

Finally, (50) is an obvious consequence of (9). In fact, since both matrices $\pm I$ are rotations, (10) implies that

$$(50 \text{ bis}) \qquad -I(\Omega_0)I,$$

which is more than (50).

31. *Proof of* (13). Criterion (i) of § 21 shows that a point O of Ω is in $\Omega - [\Omega]$ if and only if -1 is not in the spectrum of O. This implies that I is in $\Omega - [\Omega]$ and that (13) will be proved if it is shown that

$$O(\Omega - [\Omega])I \tag{51}$$

holds whenever O is an orthogonal matrix not containing -1 in its spectrum.

In order to prove (51) for every such O, apply part (β) of (IV) to $R = O$. This supplies the existence of a continuous path $R_t (0 \leqq t \leqq 1)$ which begins at the point I, ends at the point $R = O$, and is such that no matrix R_t, representing an arbitrary point of the path, will have -1 in its spectrum. In view of (i), § 21, this means that every point of the path will be in $\Omega - [\Omega]$. Since the path connects I to O, the truth of (51) follows.

32. *Proof of* (15). According to (i), § 21, and (I), an O is in the intersection of $[\Omega]$ and Ω_0 if and only if -1 is in the spectrum of O and occurs in the point spectrum of O with a multiplicity which is either ∞ or even ($\geqq 0$). Clearly, O is in $[\Omega]\Omega_0$ if and only if it is orthogonally equivalent to a matrix of the form

$$O = R \quad \text{or} \quad O = I^j \dot{+} R, \tag{52}$$

where I^j is the j-rowed unit matrix ($1 \leqq j \leqq \infty$) and R is a rotation which has -1 in its spectrum and does not have $+1$ in its point spectrum. (In the first case of (52), R is, of course, infinite; in the second case, R may be either finite or infinite.) It is sufficient to show that O can be connected by a path in $[\Omega]\Omega_0$ to a fixed matrix of this space, say to $-I$.

Let

$$R = \int_{+0}^{2\pi-0} e^{i\lambda} dE(\lambda)$$

denote the spectral resolution of R, and define R_t, for $0 \leqq t \leqq 1$, by

$$R_t = -\int_{+0}^{2\pi-0} e^{it(\lambda-\pi)} dE(\lambda).$$

Then the matrices R_t are orthogonal for all t and join $R_0 = -I^{2k}$ to $R_1 = R$; cf. (IV). It is easy to see that -1 is, for $0 < t \leqq 1$, in the spectrum of R_t, and that the multiplicity of -1 in the point spectrum of R_t is identical with that of -1 in the point spectrum of R. (In fact, if $n \to \infty$, then $(R + I)x_n \to \infty$ holds for a sequence of unit vectors x_n if and only if

$(R_t + I)x_n \to 0$ holds for every fixed t on $0 < t \leqq 1$; while $(R + I)x = 0$ holds for a vector $x \neq 0$ if and only if $(R_t + I)x = 0$ holds for every fixed t on $0 < t \leqq 1$; cf. the proof of (IV) in § 28.) Consequently, it is clear from (52) that $O([\Omega]\Omega_0)N$, where $N = -I$ or $N = M$, with

$$M = I^j \dot{+} (-I^k), \text{ where } 1 \leqq j \leqq \infty \text{ and } 1 \leqq k \leqq \infty.$$

Since the proof of (15) is complete if $N = -I$, it remains only to show that

$$M([\Omega]\Omega_0)(-I). \tag{53}$$

If $j = \infty$, then $I^j = I$, hence (53) follows, since I can be connected to $-I$ in Ω_0. If $j < \infty$, then $M = I^j \dot{+} (-I) \dot{+} (-I)$. But $I^j \dot{+} (-I)$ is a rotation and can therefore be joined to $-I$ in Ω_0, so that (53) follows as before. This completes the proof of (15).

33. Assertions (A) and (B) of § 5 will now be proved. In order to prove (A), it is sufficient to note that $Q(t) = B(\pi t) \dot{+} I$, where $0 \leqq t \leqq 1$ and where $B(\phi)$ is the binary rotation defined in § 7, is a continuous path satisfying the following conditions: $Q(1)$ is in $[\Omega]$, $Q(t)$ is in $\Omega - [\Omega]$ when $0 \leqq t < 1$, finally $I = Q(0)$. Assertion (B) is a consequence of (II); one need only choose the O of (B) to be a reflection which does not have -1 in its essential spectrum.

APPENDIX. On Bounded Matrices.

34. The following considerations concern themselves with facts which correspond to those of §21-§26 and (9)-(15) if Ω, the space of all (real) orthogonal matrices, is replaced by the larger space, say Θ, of all bounded real matrices, A. It turns out that a formal analogy results if what corresponds to the "boundary," $[\Omega]$, of Ω is taken to be the set, say $[\Theta]$, of those matrices A which fail to have a (unique) bounded reciprocal matrix, A^{-1}. Thus $\Theta - [\Theta]$ and $[\Theta]$ consist of all non-singular and of all singular matrices which represent a linear mapping of the real Hilbert space, $\mathfrak{R}$, on the whole of $\mathfrak{R}$ or on a proper subset of $\mathfrak{R}$, respectively. It is understood that Θ is meant to be the metric space on which the distance, $|A_1 - A_2|$, is defined as at the beginning of § 1.

35. What corresponds to (v) in § 22 is the following fact (which is true for finite matrices also):

(*a*) *The set* $[\Theta]$ *is closed, i. e., the set* $\Theta - [\Theta]$ *is open* (*on* Θ).

In order to prove this assertion, (α), it is sufficient to note that $\mathfrak{S}$ is a complete space and that, if A has a unique, bounded reciprocal A^{-1} (i. e., if A is in $\mathfrak{S}-[\mathfrak{S}]$), then the partial sums of both series

$$\sum_{n=0}^{\infty} A^{-1}(-A^{-1}X)^n, \qquad \sum_{n=0}^{\infty}(-XA^{-1})^n A^{-1},$$

the formal Liouville series for $(A+X)^{-1}$, form convergent sequences on $\mathfrak{S}$ whenever $|X|$ is small enough (smaller than $1/|A^{-1}|$).

Clearly, (α) implies the first of the following two assertions:

(β) *Neither* $[\mathfrak{S}]$ *nor* $\mathfrak{S}-[\mathfrak{S}]$ *is dense on* $\mathfrak{S}$.

The second assertion of (β) is false for finite matrices and represents the analogue of (iv), § 22. It will be proved by showing that there exist on $\mathfrak{S}$ points A which are not in the closure of $\mathfrak{S}-[\mathfrak{S}]$. In fact, it will be proved that an A satisfies this condition if the spectra of AA' and $A'A$ are not identical,

$$\text{(54)} \qquad \operatorname{sp}(AA') \neq \operatorname{sp}(A'A),$$

where the prime denotes the operation of transposition.

36. Condition (54) cannot of course be fulfilled by a finite matrix. That it can be satisfied by a point A of $\mathfrak{S}$, is shown by Toeplitz's example of the matrix, A, which belongs to the infinite bilinear form $x_1y_2+x_2y_3+x_3y_4+\cdots$. In fact, (54) holds for this A, since $A'A=I$ but $AA'=0+I$.

Accordingly, it is sufficient to prove the following theorem:

(V) *A point A of $\mathfrak{S}$ cannot be in the closure of $\mathfrak{S}-[\mathfrak{S}]$ if it satisfies* (54).

First, if H and H_n, where $n=1,2,\cdots$ are bounded Hermitian matrices, then, as $n\to\infty$,

$$\text{(55)} \qquad |H_n-H|\to 0 \text{ implies that } \lim \operatorname{sp}(H_n)=\operatorname{sp}(H),$$

where "$\lim \operatorname{sp}(H_n)=\operatorname{sp}(H)$" is meant to symbolize the following situation:[4] A number, say λ, is in $\operatorname{sp}(H)$ if and only if it is possible to find in every $\operatorname{sp}(H_n)$ some number, say λ_n, in such a way that $\lambda_n\to\lambda$ as $n\to\infty$.

---

[4] The truth of the implication (55), which may be well-known, follows readily in terms of spectral resolutions from the following obvious fact: If $x_1, x_2, \cdots$ denotes a sequence of unit vectors and if μ is any real number, then the limit relation $|(H-\mu I)x_n|\to 0$, as $n\to\infty$, holds if and only if $|(H_n-\mu I)x_n|\to 0$.

In this connection, cf. P. Hartman and A. Wintner, vol. 71 (1949), pp. 865-878 of this Journal.

Next, if A is in $\Theta - [\Theta]$, and if B denotes the point A^{-1} of $\Theta - [\Theta]$, then B transforms AA' into $BAA'B^{-1}$, which is $A'A$. Since the spectrum of a bounded matrix remains invariant under transformation by a bounded matrix which has a unique bounded reciprocal matrix, it follows that

$$\mathrm{sp}(AA') = \mathrm{sp}(A'A) \tag{56}$$

if A is in $\Theta - [\Theta]$.

37. In order to prove (V), § 36, suppose that it is false. Then there exists on Θ a point A satisfying (54) and having the property that $|A_n - A| \to 0$, as $n \to \infty$, holds for a sequence of points $A_1, A_2, \cdots$ contained in $\Theta - [\Theta]$. Hence, by the end of § 36,

$$\mathrm{sp}(A_n A'_n) = \mathrm{sp}(A'_n A_n)$$

holds for every n. On the other hand, $|A_n - A| \to 0$ implies that $|A'_n - A'| \to 0$ and that both product relations $|A_n A'_n - AA'| \to 0$, $|A'_n A_n - A'A| \to 0$ are true. Hence, if (55) is applied to both product sequences, it follows that

$$\lim \mathrm{sp}(A_n A'_n) = \mathrm{sp}(AA'), \qquad \lim \mathrm{sp}(A'_n A_n) = \mathrm{sp}(A'A).$$

But the last two formula lines imply (56) for the given A. The latter was, however, supposed to satisfy (54). This contradiction proves (V), § 36.

38. The result can be expressed as follows: Condition (56) is necessary[5] in order that a point A of Θ be in the closure of $\Theta - [\Theta]$.

It is known that, if A is any finite matrix, then AA' and $A'A$ are orthogonally equivalent, which is more than (56). One might therefore expect that, for a point A of $\Theta - [\Theta]$, the necessary condition (56) can be improved to the orthogonal equivalence of AA' and $A'A$. But this refinement of (56) proves to be false. A counter-example is supplied by the matrix, A, of the bilinear form $\epsilon_1 x_1 y_2 + \epsilon_2 x_2 y_3 + \epsilon_3 x_3 y_4 + \cdots$, if $\epsilon_1, \epsilon_2, \cdots$ is any sequence of non-vanishing numbers satisfying $\epsilon_n \to 0$ as $n \to \infty$.

In fact, the latter condition implies that A is completely continuous. This in turn implies (cf. below) that A is in the closure of $\Theta - [\Theta]$. Nevertheless, AA' and $A'A$ are not orthogonally equivalent, since they are two diagonal matrices one of which does, while the other does not, contain 0 in the diagonal, since $\epsilon_n^2 \neq 0$.

[5] That condition (56) is not sufficient as well is shown by the simple example, pointed out to us by Professor Hartman, which results if the example of Toeplitz, referred to in § 36, is bordered by a row of zeros and a column of zeros.

A class substantially more general than that of all real completely continuous matrices is defined by the following requirement for a point A of $\mathfrak{O}$: There does not exist any positive $a = a(A)$ having the property that $(\lambda I - A)^{-1}$ fails to exist (as a unique, bounded reciprocal) for every λ satisfying $|\lambda| < a$. Then $\lambda_n I - A$ is in $\mathfrak{O} - [\mathfrak{O}]$ for certain $\lambda_n \rightarrow 0$ as $n \rightarrow \infty$. Since this implies that $A_n = A - \lambda_n I$ satisfies $|A_n - A| \rightarrow 0$ on $\mathfrak{O}$, it follows that A is in the closure of $\mathfrak{O} - [\mathfrak{O}]$ (hence, in particular, (56) is satisfied). Cf. [3], p. 241.

Another sufficient condition in order that a point A of $\mathfrak{O}$ be in the closure of $\mathfrak{O} - [\mathfrak{O}]$ consists in $AA' = A'A$. In fact, A is then normal, hence such as to have a spectral matrix, from which the existence of a sequence $A_1, A_2, \cdots$ contained in $\mathfrak{O} - [\mathfrak{O}]$ and satisfying $|A_n - A| \rightarrow 0$ can readily be concluded.

39. Let Γ be the metric space of *all* (real or *complex*) bounded matrices A, the distance between two points, A_1 and A_2, of Γ being defined as the least upper bound of the length of the vector $(A_1 - A_2)y$ when y varies over all unit vectors of the complex Hilbert space, and let $[\Gamma]$ denote the subset of Γ consisting of those matrices A for which there does not exist a (unique) bounded reciprocal, A^{-1}. Then it is clear from the above proofs that theorems

(α), (β) *of* § 35 *and* (V) *of* § 36 *remain true if* $\mathfrak{O}$, $[\mathfrak{O}]$ *and* (54) *are replaced by* Γ, $[\Gamma]$ *and* $\mathrm{sp}(AA^*) \neq \mathrm{sp}(A^*A)$, *respectively,*

where A^* denote the complex conjugate of A'.

THE INSTITUTE FOR ADVANCED STUDY
AND
THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] M. H. Martin, "On infinite orthogonal matrices," *American Journal of Mathematics*, vol. 54 (1932), pp. 579-631.

[2] C. R. Putnam and A. Wintner, "The connectedness of the orthogonal group in Hilbert space," *Proceedings of the National Academy of Sciences*, vol. 37 (1951), pp. 110-112.

[3] A. Wintner, "Zur Theorie der beschränkten Bilinearformen," *Mathematische Zeitschrift*, vol. 30 (1929), pp. 228-282.

[4] ———, *Spektraltheorie der unendlichen Matrizen*, Leipzig, 1929.

ON PERTURBATIONS OF THE CONTINUOUS SPECTRUM OF THE HARMONIC OSCILLATOR.*

By Philip Hartman and Aurel Wintner.

1. Let $f(t)$ be a real-valued continuous function on the half-line $0 \leqq t < \infty$. Consider on the latter the differential equation

$$x'' + (\lambda + f)x = 0 \tag{1}$$

with a homogeneous, linear boundary condition at $t = 0$, such as $x(0) = 0$ or, more generally,

$$x(0) \cos \alpha + x'(0) \sin \alpha = 0, \tag{2}$$

where $0 \leqq \alpha < \pi$. If (1) is of *Grenzpunkt* type, that is, if (1) and (2) determine a self-adjoint problem on the $L^2(0, \infty)$-space (for some and/or every α), let S_α denote the spectrum of this problem. Finally, let C_α denote that subset of S_α which represents the continuous spectrum.

The latter will be meant in Hilbert's sense, that is, in terms of the λ-set of non-constancy of the sum of the two continuous components (if any) of the spectral resolution analyzed into its three Lebesgue components, which are absolutely continuous, continuous but purely singular and purely discontinuous, respectively. No example of a self-adjoint problem (1)-(2) seems to be known in which the second of these three components is present.

In the case of the harmonic oscillator

$$x'' + \lambda x = 0, \tag{3}$$

that is, in the case $f(t) \equiv 0$ of (1), the explicit form of the general solution shows that

(4) (1) is of *Grenzpunkt* type

and that every non-negative λ is in C_α, while no negative λ is in the closure of S_α; so that, in particular,

$$S_\alpha' = [0, \infty) \tag{5}$$

where S_α' is the closure of S_α. In what follows, there will be delimited for an arbitrary "perturbation," $f(t)$, of (3) that "degree of smallness" (for

* Received November 16, 1949; revised October 4, 1951.

large t) under which (4) and at least the following weakened form of (5) prove to be true:

$$S_\alpha' \supset [0, \infty) \text{ in } (1). \tag{6}$$

It should be mentioned that the closure, S_α', of S_α, the so-called essential spectrum of (1), is always independent of α (whenever (1) holds); cf. [7], p. 251 and, for direct characterizations of this α-invariant λ-set, [3], [3 bis].

2. A sufficient "degree of smallness" turns out to be the following specification of a "small" f:

$$\int_0^T |f(t)|\, dt = o(T) \tag{7}$$

or, more generally,

$$\liminf_{T\to\infty} T^{-1} \int_0^T |f(t)|\, dt = 0. \tag{8}$$

In other words, (8) implies both (4) and (6). This criterion proves to be of a final nature, in the sense that (8) cannot be relaxed to

$$\liminf_{T\to\infty} T^{-1} \int_0^T |f(t)|\, dt \leqq c < \infty, \tag{9}$$

nor even to

$$\limsup_{T\to\infty} T^{-1} \int_0^T |f(t)| dt < \infty, \tag{10}$$

since (10) is insufficient for (6); cf. the end of Section **7**. On the other hand, it can be concluded from [4] that (10), and even (9), is sufficient for (4) alone.

3. It is natural to ask why to consider (6), instead of the sharper statement (5). The answer is that (5) can fail to be true if (8), or for that matter (7), is satisfied. In fact, it will be shown in Section **6** that not only the negation of (5) but even *the possibility of*

$$S_\alpha' = (-\infty, \infty) \tag{11}$$

can be realized if only (7) *is assumed.* Thus, while (6) can,

$$S_\alpha' \subset [0, \infty) \tag{12}$$

cannot, be concluded from (8) or (7).

The situation is changed entirely if (7) is replaced by the more drastic assumption

$$f(t) = o(1), \tag{13}$$

where $t \to \infty$. In fact, it was shown in [2] that (13) implies (5). But the proofs of the implication (13) $\to$ (5) in [2] and [5] are quite involved. A simple proof follows from the general result in [3]. It will be shown under (i) in Section **7** below that (13) $\to$ (5) can also be proved very simply by an appeal to standard theorems in the theory of operators in Hilbert space.

4. As mentioned above, both (8) $\to$ (4) and (8) $\to$ (6) are true but in the second of these conclusions, the hypothesis (8) cannot be relaxed to (10). But it turns out that (8) $\to$ (6) can be improved in a direction which relaxes the hypothesis (8) to (9), with a *fixed* $c \geqq 0$ and in such a way that the conclusion (6) becomes replaced by one depending on the numerical value of c and leading to (6) when $c = 0$.

In order to formulate this refinement of (8) $\to$ (6), it will be convenient to use the following definition: With reference to a fixed positive number c, a differential equation (1) has the property (c) if (4) is satisfied and if *every* interval of the form

$$[\lambda, \lambda + 4c + 4c^2/\lambda], \text{ where } \lambda > 0, \tag{14}$$

contains at least one point of S_a'. Then the generalization in question can be formulated as follows: *For any given* $c \geqq 0$, *property* (c) *holds whenever* (9) *is satisfied.* This will be seen at the end of Section **5**.

Except for the upper end point of the interval in (14), which is sharper in [5], the last italicized statement is a generalization of a result, proved in [5], which claims the corresponding property (c) under the assumption that

$$|f(t)| \leqq c, \tag{15}$$

rather than just (9), is satisfied.

5. The proofs will be based on the following fact: If $f(t)$ is real-valued and continuous, and if $N(T, \lambda)$ denotes the number of zeros of a (real-valued) solution $x(t) = x_\lambda(t) \not\equiv 0$ of (1) on the interval $0 \leqq t \leqq T$, then the inequality

$$|\pi N(T, \lambda) - \lambda^{\frac{1}{2}} T| \leqq 2\pi + \lambda^{-\frac{1}{2}} \int_0^T |f(t)|\, dt \tag{16}$$

holds for every $\lambda > 0$ and for every $T > 0$.

In order to see this, use will be made of the fact that, since $x(t)$ and $x'(t)$ cannot vanish simultaneously, the relation

$$\theta(t) = \arctan \{\lambda^{\frac{1}{2}} x(t)/x'(t)\} \tag{17}$$

and the choice of an initial determination $(\operatorname{mod} \pi)$ at $t = 0$ (such as $0 \leqq \theta(0) < \pi$) define a unique *continuous* function $\theta = \theta(t)$, and that the latter has a derivative $\theta'(t)$, given by

$$\theta' = \lambda^{\frac{1}{2}} + f\lambda^{\frac{1}{2}} x^2/(\lambda x^2 + x'^2). \tag{18}$$

This follows from (17) by virtue of (1).

If $\lambda > 0$, it follows from (18) that

$$|\theta'(t) - \lambda^{\frac{1}{2}}| \leqq \lambda^{-\frac{1}{2}} |f(t)|. \tag{19}$$

On the other hand, it is seen from (17) that $\theta(t)$ becomes an integral multiple of π at exactly those t-values which are zeros of $x(t)$, while (18) shows that $\theta'(t) > 0$ whenever $x(t) = 0$. Hence it is seen from the definition of $N(T, \lambda)$ that

$$|\pi N(T, \lambda) - \theta(T) + \theta(0)| \leqq 2\pi. \tag{20}$$

In order to obtain (16), it is sufficient to combine (20) with the inequality which results if (19) is integrated between $t = 0$ and $t = T$.

Proof of (8) $\rightarrow$ (4). It follows from (8) and (16) that

$$\liminf_{T \to \infty} N(T, \lambda)/T \tag{21}$$

has the value $\lambda^{\frac{1}{2}}/\pi$. In particular, (21) is distinct from ∞. According to criterion (**) of [4], p. 207, this is sufficient for (4).

Proof of (8) $\rightarrow$ (6). Let $0 < \lambda < \mu$. Then, by (16),

$$|\pi[N(T, \mu) - N(T, \lambda)] - (\mu^{\frac{1}{2}} - \lambda^{\frac{1}{2}})T| \leqq 4\pi + 2\lambda^{-\frac{1}{2}} \int_0^T |f(t)|\, dt. \tag{22}$$

It follows therefore from (8) that

$$\limsup_{T \to \infty} [N(T, \mu) - N(T, \lambda)] = \infty. \tag{23}$$

But Sturm's separation theorem implies that the inequality

$$N(t, \mu) - N(t, \lambda) \geqq N(T, \mu) - N(T, \lambda) - 1$$

holds for every $t > T$. Hence it is clear that the "lim sup" in (23) can be replaced by a "lim," i. e., that

$$N(T,\mu) - N(T,\lambda) \to \infty \text{ as } T \to \infty. \tag{24}$$

Finally, it was shown in [3], p. 915, that the latter condition, (24), is sufficient in order that at least one point of S_α' be contained in the interval $[\lambda, \mu]$. Since (24) holds whenever $0 < \lambda < \mu$, the proof of (6) is complete.

Proof of $(9) \to (c)$. This implication (i. e., the italicized assertion of Section **4**) follows from (22) in the same way as $(8) \to (6)$ did. In fact, (9) implies (23) if $\mu^{\frac12} - \lambda^{\frac12} \geqq 2c/\lambda^{\frac12}$, that is, if $\mu > \lambda + 4c + 4c^2/\lambda$.

6. The italicized statement of Section **3** concerning (11) will be proved by an example.* To this end, let $f(t)$ be a function which is continuous for $0 \leqq t < \infty$, satisfies (7) and has the constant value m on the interval

$$m^4 \leqq t \leqq m^4 + m, \tag{25}$$

where $m = 1, 2, \cdots$. Such functions exist, since the contribution of the intervals (25) to the integral occurring in (7) is $o(T)$, simply because

$$\sum_{m=1}^{M} \int_{m^4}^{m^4+m} |f(t)|\, dt = \sum_{m=1}^{M} m^2 = o(M^4).$$

Let λ be any point of the line $-\infty < \lambda < \infty$. Since $f(t) \equiv m$ on the interval (25), a solution of (1) on (25) is $x(t) = \cos\,(m+\lambda)^{\frac12} t$ whenever m is so large that $m + \lambda > 0$. Consider only such values of m. Then, by Sturm's comparison theorem, the contribution of the interval (25) to the number $N(T,\lambda)$ is between the bounds

$$\{(m+\lambda)^{\frac12} m - 2\pi\}/\pi \text{ and } \{(m+\lambda)^{\frac12} + 2\pi\}/\pi$$

whenever $T \geqq m^4 + m$. Hence, if $\mu > \lambda$, the contribution of such an interval (25) to the difference in (24) is

$$\geqq \{[(m+\mu)^{\frac12} - (m+\lambda)^{\frac12}]m - 4\pi\}/\pi.$$

Since the last $\{\ \ \}$ is $\sim \frac12(\mu - \lambda)m^{\frac12}$ as $m \to \infty$, it follows that (24) is satisfied. Hence, if (I) in [3], p. 915, is applied in the same way as at the end of the Proof of $(8) \to (6)$ (Section **5**), then, since $[\lambda, \mu]$ is now *any* interval, (11) follows.

---

* In a similar context, a corresponding example was communicated to us by Dr. C. R. Putnam.

7. The following remarks, (i) and (ii), do not contain new results, but simple proofs of known theorems.

(i) Suppose that $f(t)$ satisfies (13). Then the difference of the two self-adjoint operators on $L^2(0,\infty)$ which are defined by a common boundary condition (2) and by the differential operators corresponding to (1) and (3), respectively, is completely continuous. In fact, the difference of these two self-adjoint operators is represented by the operator $f(t)x(t)$ which, in view of (13) (and of the continuity of f), is completely continuous on the x-space $L^2(0,\infty)$. But (5) is true in the case (3). It follows therefore from a classical theorem ([6], p. 384; cf. [1], p. 120) that (5) holds in the case (13) of (1) also.

(ii) Besides (1), consider another differential equation of the form (1), say

$$x'' + (\lambda + g)x = 0, \tag{26}$$

where $f(t)$ and $g(t)$ are real-valued and continuous for $0 \leqq t < \infty$. It is well-known that, if

$$| g(t) - f(t) | < \text{const.}, \qquad (0 \leqq t < \infty), \tag{27}$$

then (1) must be of *Grenzpunkt* type whenever (26) is. In fact, this can be refined to an explicit asymptotic connection (as $t \to \infty$) between the general solution of (26) and that of (1), without assuming more than (27) and the negation of (4); see [8]. If this refinement is not required, a simple proof can be obtained along the lines of the argument applied under (i) above. In fact, what corresponds to the deviation $f(t)x(t)$ in (i) is now the difference

$$\{f(t) - g(t)\}x(t); \tag{28}$$

cf. (1) and (26). But (27) assures that (28) is a bounded operator on the x-space $L^2(0,\infty)$. Hence it is sufficient to apply a classical theorem ([1], p. 78), according to which the sum of a self-adjoint and of a bounded operator is always self-adjoint.

The proof in (i) was based on a general criterion concerning the perturbation of a self-adjoint operator by a completely continuous one. It may be mentioned that this criterion can be generalized to a theorem on the perturbation of the spectrum of a self-adjoint operator by the addition of an arbitrary bounded self-adjoint operator (having a norm not exceeding a given value, $c > 0$). The general theorem in question implies, as a special case, the result of [5] concerning the spectra of differential operators associated with (1) and (26), subject to the condition (27).

If (ii) is applied to $g = 0$, then, since (1), (26) reduce to (1), (3), respectively, it follows that (15) is sufficient for (4). On the other hand, (6) does not follow from (15), or even from (15) and

$$(29) \qquad \liminf_{T \to \infty} | f(t) | = 0$$

together. This is seen by choosing $f(t) = a \cos t$, where a is any non-vanishing constant. In fact, (1) then becomes Mathieu's equation, for which the positive λ-values not contained in S_a' are known to form a sequence of intervals ("instability regions") which cluster at $\lambda = \infty$.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] T. Carleman, *Sur les équations intégrales singulières á noyau réel et symétrique*, Uppsala, 1923.

[2] P. Hartman, "On the spectra of slightly disturbed linear oscillators," *American Journal of Mathematics*, vol. 71 (1949), pp. 71-79.

[3] ———, "A characterization of the spectra of one-dimensional wave equations," *ibid.*, pp. 915-920.

[3 bis] ——— and A. Wintner, "On the essential spectra of singular eigenvalue problems," *ibid.*, vol. 72 (1950), pp. 545-564.

[4] ——— and A. Wintner, "A criterion for the non-degeneracy of the wave equations," *ibid.*, pp. 206-213.

[5] C. R. Putnam, "The cluster spectra of bounded potentials," *ibid.*, pp. 612-620.

[6] H. Weyl, "Ueber beschränkte quadratische Formen, deren Differenz vollstetig ist," *Rendiconti del Circolo Matematico di Palermo*, vol. 27 (1909), pp. 373-392.

[7] ———, "Ueber gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen," *Mathematische Annalen*, vol. 68 (1910), pp. 220-269.

[8] A. Wintner, "Asymptotic integrations of the adiabatic oscillator," *American Journal of Mathematics*, vol. 69 (1947), pp. 251-272.

ON THE GENUS OF THE FUNDAMENTAL REGION OF SOME SUBGROUPS OF THE MODULAR GROUP.*

By EMIL GROSSWALD.

1. The purpose of this paper is to use uniformly a general method, in order to establish the genus of the fundamental region for congruence subgroups (modulo primes) of the modular group. The results for the subgroups $\Gamma(p)$ and $\Gamma_0{}^0(p)$ are known (see [2], p. 249, and [4], p. 832, respectively); those for $\Gamma_0(p)$ are believed to be new. Let G be a group of transformations, R a simply connected fundamental region for an automorphic function admitting the group G. Let $2Q$ be the number of sides of R, C the number of cycles of its corners and P the genus of R. Then

$$P = \tfrac{1}{2}(Q - C + 1). \tag{1}$$

(see [1], p. 239, also [3, I], p. 262, (2)). Let m be the total number of independent generators and ρ the number of parabolic generators of G; let ν be the number of independent defining relations satisfied by the generators of G. Furthermore, let n be the number of cycles of R, corresponding to fixed points, μ the number of remaining cycles (corresponding to "accidental" corners). Then $n = \nu + \rho$ and the following relations hold (see [3, I], p. 170 and p. 262):

(2) $\qquad Q = M; \qquad\qquad$ (3) $\qquad C = n + \mu = \nu + \rho + \mu.$

2. Let Γ be the group of nonhomogeneous modular transformations. Its elements are the matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, where the integers a, b, c, d satisfy $ad - bc = 1$ and where we consider $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and $\begin{pmatrix} -a & -b \\ -c & -d \end{pmatrix}$ as identical elements. Γ admits the system of generators $S = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $T = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$.

3. For every prime number p we define the subgroups $\Gamma_0(p)$ of Γ by the additional condition $c \equiv 0 \pmod p$. These subgroups have been studied by H. Rademacher [5], who showed that, if the square brackets stand for

* Received June 14, 1950; in revised form May 23, 1951.

the greatest integer function and the round brackets are Legendre symbols, then, for $p > 3$,

i) $\Gamma_0(p)$ is generated by $m = 2[p/12] + 3$ independent generators, satisfying $\nu = 2 + (-1/p) + (-3/p)$ defining relations;

ii) one of the generators of $\Gamma_0(p)$ is S^p, the others are of the form $V_k = \begin{pmatrix} -k_* & -1 \\ kk_* + 1 & k \end{pmatrix}$, with

$$kk_* + 1 \equiv 0 \pmod p, \qquad 1 < k, k_* < p - 1. \tag{4}$$

From (4) follows immediately the following

LEMMA. *For every p, $\Gamma_0(p)$ has exactly one parabolic generator.*

Proof. S^p is parabolic for every p. For $p = 2$ and $p = 3$, we verify directly that $V_1 = \begin{pmatrix} -1 & -1 \\ 2 & 1 \end{pmatrix}$, $V_2 = \begin{pmatrix} -1 & -1 \\ 3 & 2 \end{pmatrix}$, respectively, are not parabolic. If $p > 3$ and V_k is parabolic, it follows (see [1], p. 21) successively, by (4), that $k - k_* = \pm 2$, or $k^2 - kk_* = \pm 2k$, so that $(k \pm 1)^2 \equiv 0 \pmod p$, contrary to (4), proving the lemma.

With the indicated values of m and ν, it follows from (2), (3) and the lemma, that $Q = 2[p/12] + 3$ and $n = \{2 + (-1/p) + (-3/p)\} + 1$. In the case of groups of modular transformations, $\mu = 1$ (see [3, I], p. 262 and 319) so that (1) becomes $P = [p/12] - \frac{1}{2}\{(-1/p) + (-3/p)\}$. Observing that the large bracket is a periodic function (mod 12) of p, we can write the last relation also as $P = [p/12] + r(r^2 - 25)/24$, where $r \equiv p \pmod{12}$, $|r| \leqq 5$.

4. The subgroups $\Gamma_0{}^0(p)$ are defined, by adding to the previous condition $c \equiv 0 \pmod p$, the new condition $b \equiv 0 \pmod p$. For $p > 3$, the values of m, ν and ρ are respectively (see [4]), $m = 2[(p+2)(p-1)/12] + 3$, $\nu = 2 + (-1/p) + (-3/p)$ and $\rho = 3$. Substituting these values in (2) and (3) we obtain from (1), $P = [(p+2)(p-1)/12] - 1 + r(r^2 - 25)/24$, where r is defined as before.

5. The principal subgroups $\Gamma(p)$ are defined, by adding to the definition of $\Gamma_0{}^0(p)$ the further conditions $a \equiv d \equiv 1 \pmod p$. Their structure has been studied by H. Frasch [2] and J. Nielsen [6]. H. Frasch showed that, for $p > 3$, $\Gamma(p)$ is generated as a free group by $m = p(p^2 - 1)/12 + 1$ independent generators; one of them is S^p, the others depend on three parameters. The fundamental region R', to which Frasch's generators

correspond, is not simply connected; therefore, we cannot apply (1), (2) and (3) directly, but have first to transform R' into a simply connected fundamental region R. In this transformation, the number of sides does not change, so that $Q = m = p(p^2-1)/12+1$. Furthermore, $\Gamma(p)$ being a free group, it contains no elements of finite order. Consequently, the only cycles of R are those corresponding to transforms of S. As the index of $\Gamma(p)$ in Γ is $j = \frac{1}{2}p(p^2-1)$ and $S^h \varepsilon \Gamma(p)$ if, and only if $p \mid h$, it follows (see [2], p. 248) that $C = j/p = \frac{1}{2}(p^2-1)$. From (1) follows now the (well known) value

$$P = \tfrac{1}{2}\{p(p^2-1)/12+1-\tfrac{1}{2}(p^2-1)+1\} = (p^2-1)(p-6)/24+1.$$

6. So far we have considered only the cases $p > 3$. If $p = 2$, or 3, a direct examination yields the well-known result, that the genera of $\Gamma_0(p)$, $\Gamma_0{}^0(p)$ and $\Gamma(p)$ are all zero.

UNIVERSITY OF SASKATCHEWAN.

BIBLIOGRAPHY.

[1] L. R. Ford, *Automorphic Functions*, New York, 1929.

[2] H. Frasch, "Die Erzeugenden der Hauptkongruenzgruppen für Primzahlstufen," *Mathematische Annalen*, vol. 108 (1933), pp. 229-252.

[3] R. Fricke-F. Klein, *Vorlesungen über die Theorie der automorphen Functionen*, vol. I (Leipzig, 1897) and vol. II (Leipzig, Berlin, 1912).

[4] E. Grosswald, "On the structure of some subgroups of the modular group," *American Journal of Mathematics*, vol. 72 (1950), pp. 809-834.

[5] H. Rademacher, "Ueber die Erzeugenden von Kongruenzuntergruppen der Modulgruppe," *Abhandlungen aus dem Seminar der Hamburgischen Universität*, vol. 7 (1929), pp. 134-138.

[6] J. Nielsen, "A study concerning the congruence subgroups of the modular group," *Det Kgl, Danske Videnskabernes Selskab Matematisk-Fysiske Meddelelsen*, vol. 25, no. 18 (1950).

CYCLOTOMY AND JACOBSTHAL SUMS.*

By ALBERT LEON WHITEMAN.

1. Introduction. Let p be an odd prime and e be a divisor of $p-1$. The Jacobsthal sum $\phi_e(n)$ is defined by

$$\phi_e(n) = \sum_{h=1}^{p-1} \left(\frac{h}{p}\right)\left(\frac{h^e+n}{p}\right), \tag{1.1}$$

where the symbol (h/p) denotes the quadratic character of h with respect to p. In [5] Jacobsthal studied $\phi_2(n)$ and deduced an interesting application in connection with the representation of a prime p of the form $4m+1$ as the sum of two squares. Later writers [2], [6], [7], [9] obtained analogous results for other values of e. These results exhibit a connection between the Jacobsthal sum and a partition of p into quadratic summands. This connection is made precise in Theorem 1 of the present paper (§ 4). The first part of the paper (§§ 2-4) contains an account of the arithmetic properties of the sum $\phi_e(n)$ and the related sum $\psi_e(n)$ defined in § 3. There is given in § 5 a result which expresses the Jacobsthal sum in terms of a certain cyclotomic function. Finally the remainder of the paper (§§ 6-9) contains numerous applications.

2. The Jacobsthal sum. Let g be a fixed primitive root of p and write $p-1=ef$. Assume first that n in (1.1) is divisible by p. Then we see that $\phi_e(n)=0$ or $p-1$ according as e is even or odd. Next let n be prime to p, so that n is congruent to a power g^m, $m=0,1,\cdots, p-2$, of the primitive root g. From the definition it is clear that the value of $\phi_e(g^m)$ depends upon the primitive root g employed except when $m=0$. We now show that $\phi^2_e(g^m)=\phi^2_e(g^{m'})$ if $m\equiv m' \pmod e$. This result is included in the formula

$$\phi_e(g^{et+k}) = (-1)^{t(e+1)}\phi_e(g^k), \tag{2.1}$$

where $0\leqq t\leqq f-1$, $0\leqq k\leqq e-1$. To prove (2.1) let $\bar{h}$ denote any solution of the congruence $h\bar{h}\equiv 1 \pmod p$. Then we have

$$\sum_{h=1}^{p-1}\left(\frac{h}{p}\right)\left(\frac{h^e+g^{et+k}}{p}\right)=\left(\frac{g^t}{p}\right)^{e+1}\sum_{h=1}^{p-1}\left(\frac{\bar{h}g^t}{p}\right)^{e+1}\left(\frac{1+(\bar{h}g^t)^e g^k}{p}\right)$$
$$=(-1)^{t(e+1)}\sum_{h=1}^{p-1}\left(\frac{h}{p}\right)\left(\frac{h^e+g^k}{p}\right).$$

* Received July 18, 1950.

A formula closely related to (2.1) is

$$\phi_e(nx^e) = (x/p)^{e+1}\phi_e(n) \qquad (p \nmid x). \tag{2.2}$$

This may be directly established.

Suppose now that f is odd. Then e must be even. Since e is a divisor of $p-1$, we may select x in (2.2) so that x belongs to the exponent e (mod p). Then employing Euler's criterion, we get $(x/p) \equiv x^{(p-1)/2} = (x^{e/2})^f \equiv (-1)^f \equiv -1 \pmod p$. Hence in this case $\phi_e(n) = -\phi_e(n)$. We have therefore proved

$$\phi_e(n) = 0 \qquad (f \text{ odd}). \tag{2.3}$$

We next prove that

$$\phi_e(n) \equiv 0 \pmod e \qquad (f \text{ even}). \tag{2.4}$$

The congruence $h^e \equiv 1 \pmod p$ has e incongruent roots since e is a divisor of $p-1$. Let r belong to the exponent e (mod p). Then for a fixed value of x not divisible by p, the e roots of the congruence $h^e \equiv x^e \pmod p$ are given by $h_i \equiv r^i x \pmod p$, $i = 0, 1, \cdots, e-1$. Now for f even we have $(r/p) \equiv r^{(p-1)/2} = (r^e)^{f/2} \equiv 1 \pmod p$. Hence $(h_i/p) = (x/p)$. We may now deduce (2.4) at once from (1.1).

Finally we establish the congruence

$$\phi_e(n) \equiv -\sum_{j=0}^{[(e-1)/2]} \binom{(p-1)/2}{(2j+1)(p-1)/2e} n^{(p-1)(e-2j-1)/2e} \qquad \pmod p \; (f \text{ even}). \tag{2.5}$$

Since f is even, $p-1$ is divisible by $2e$. Using (1.1) and Euler's criterion we expand $(h^{e+1} + nh)^{(p-1)/2}$ by the binomial theorem and interchange signs of summation; the result is

$$\phi_e(n) \equiv \sum_{v=0}^{(p-1)/2} \binom{(p-1)/2}{v} n^{(p-1)/2-v} \sum_{h=1}^{p-1} h^{(p-1)/2+ev} \qquad \pmod p.$$

In order to complete the proof we make use of the formula

$$\sum_{h=1}^{p-1} h^s \equiv \begin{cases} -1 \pmod p & (s \equiv 0 \pmod{p-1}), \\ 0 \pmod p & (s \not\equiv 0 \pmod{p-1}), \end{cases}$$

and note that $(p-1)/2 + ev$ is divisible by $p-1$ if and only if ev is an odd multiple of $(p-1)/2$.

3. The sum $\Psi_e(n)$. Related to the Jacobsthal sum $\phi_e(n)$ is the sum $\psi_e(n)$ defined as follows:

$$\psi_e(n) = \sum_{h=1}^{p-1} \left(\frac{h^e + n}{p}\right). \tag{3.1}$$

If n is divisible by p we have immediately $\psi_e(n) = p - 1$ or 0 according as e is even or odd.

For $e = 2$ we have the following formula

$$\psi_2(n) = p - 1 - (n/p) - (n/p)^2 p, \tag{3.2}$$

where the symbol (n/p) is defined as 0 when n is divisible by p. Formula (3.2) is well-known and may be proved without much difficulty. Note that $\psi_2(n)$ is equal to $p-1$ when n is divisible by p and is equal to $-1-(n/p)$ otherwise.

In the sequel we shall encounter the sum

$$f(a, b, c) = \sum_{x=1}^{p-1} \left(\frac{ax^2 + bx + c}{p}\right) \qquad (p \nmid a), \tag{3.3}$$

for which we have the relation

$$f(a, b, c) = -(c/p) + (-aD/p) + (a/p)\psi_2(-D) \qquad (D = b^2 - 4ac). \tag{3.4}$$

The following two formulas are analogous to (2.1) and (2.2) and are proved in much the same way.

$$\psi_e(g^{et+k}) = (-1)^{te}\psi_e(g^k) \qquad (0 \leqq t \leqq f-1, 0 \leqq k \leqq e-1). \tag{3.5}$$

$$\psi_e(nx^e) = (x/p)^e\psi_e(n) \qquad (p \nmid x). \tag{3.6}$$

We may also establish the following formulas:

$$\phi_e(g^k) = \begin{cases} (-1)^{k+1}\phi_e(g^{e-k}) & (e \text{ even}), \\ (-1)^{k+1}\psi_e(g^{e-k}) & (e \text{ odd}), \end{cases} \tag{3.7}$$

$$\psi_e(g^k) = (-1)^k\psi_e(g^{e-k}) \qquad (e \text{ even}), \tag{3.8}$$

for $0 \leqq k \leqq e-1$. For example, (3.7) follows from

$$\sum_{h=1}^{p-1}\left(\frac{h}{p}\right)\left(\frac{h^e + g^k}{p}\right) = \left(\frac{g^{e-k}}{p}\right)\left(\frac{g}{p}\right)^{e+1}\sum_{k=1}^{p-1}\left(\frac{\bar{h}g}{p}\right)^{e+1}\left(\frac{(\bar{h}g)^e + g^{e-k}}{p}\right).$$

A formula of a different nature is

$$\phi_e(n) + \psi_e(n) = \psi_{2e}(n) \qquad (f \text{ even}). \tag{3.9}$$

This may be proved by summing in (1.1) and (3.1) first with respect to the quadratic residues of p and then with respect to the quadratic non-residues. Note that since f is even, $p-1$ is divisible by $2e$.

Finally we shall prove the following congruences for $0 \leqq k \leqq e-1$.

$$(3.10) \qquad \psi_e(g^k) \equiv \begin{cases} e \pmod{2e} & (f \text{ even}, k = 0 \text{ or } f \text{ odd}, k \neq e/2), \\ 0 \pmod{2e} & (f \text{ even}, k \neq 0 \text{ or } f \text{ odd}, k = e/2). \end{cases}$$

To establish (3.10) we write $\psi_e(g^k)$ in the form $eA - eB$, where eA is the number of times that the symbol $((h^e + g^k)/p)$ takes on the value 1 and eB is the number of times that it takes on the value -1. On the other hand $eA + eB$ is equal to $p - 1 - e$ or $p - 1$ according as the congruence $h^e + g^k \equiv 0 \pmod p$, $1 \leqq h \leqq p - 1$, does or does not have solutions. Since $g^{(ef)/2} = g^{(p-1)/2} \equiv -1 \pmod p$, we may write this congruence in the form $h^e \equiv g^{(fe)/2+k} \bmod p)$. For f even the last congruence has solutions if and only if $k = 0$; for f odd it has solutions if and only if $k = e/2$. Eliminating B by addition we get for $0 \leqq k \leqq e - 1$

$$(3.11) \quad \psi_e(g^k) = \begin{cases} -p + 1 + e + 2eA & (f \text{ even}, k = 0 \text{ or } f \text{ odd}, k = e/2), \\ -p + 1 + 2eA & (f \text{ even}, k \neq 0 \text{ or } f \text{ odd}, k \neq e/2), \end{cases}$$

from which (3.10) may easily be derived.

4. The main results. We first prove

$$(4.1) \qquad \sum_{k=0}^{e-1} \phi_e(g^k) = -e \qquad (e \text{ odd}).$$

To prove this we make use of (2.1) with e odd. Thus we get

$$\frac{p-1}{e} \sum_{k=0}^{e-1} \phi_e(g^k) = \sum_{t=0}^{f-1} \sum_{k=0}^{e-1} \phi_e(g^{et+k}) = \sum_{m=0}^{p-2} \phi_e(g^m)$$
$$= -(p-1) + \sum_{a=0}^{p-1} \phi_e(a) = -(p-1).$$

Equation (4.1) follows immediately.

We now state our principal result.

THEOREM 1. *Let s be a fixed integer such that $0 \leqq s \leqq e - 1$. If e is odd then*

$$(4.2) \qquad \sum_{k=0}^{e-1} \phi_e(g^k)\phi_e(g^{k+s}) = \begin{cases} e^2 p - e(p-1) & (s = 0), \\ -e(p-1) & (s \neq 0). \end{cases}$$

If e is even and f is even then

$$(4.3) \qquad \sum_{k=0}^{e-1} \phi_e(g^k)\phi_e(g^{k+s}) = \begin{cases} e^2 p & (s = 0), \\ 0 & (s \neq 0). \end{cases}$$

To prove this theorem let S denote the left member of (4.2) or (4.3).

It should be noted that when f is odd each term of S is equal to zero in view of (2.3). We get by (1.1) and (2.1)

$$\begin{aligned}\frac{p-1}{e}S &= \sum_{t=0}^{f-1}\sum_{k=0}^{e-1}\phi_e(g^{et+k})\phi_e(g^{et+k+s}) \\ &= \sum_{m=0}^{p-2}\phi_e(g^m)\phi_e(g^{m+s}) \\ &= \sum_{x,y=1}^{p-1}\left(\frac{x}{p}\right)\left(\frac{y}{p}\right)\sum_{m=0}^{p-2}\left(\frac{g^s g^{2m}+(x^e g^s+y^e)g^m+x^e y^e}{p}\right) \\ &= \sum_{x,y=1}^{p-1}\left(\frac{x}{p}\right)\left(\frac{y}{p}\right)f(g^s, x^e g^s+y^e, x^e y^e),\end{aligned}$$

where $f(a,b,c)$ is defined in (3.3). Making use of (3.2) and (3.4) this becomes

$$\frac{p-1}{e}S = -\sum_{x,y=1}^{p-1}\left(\frac{xy}{p}\right)^{e+1} - \left(\frac{g^s}{p}\right)p\sum_{x=1}^{p-1}\left(\frac{x}{p}\right)\sum_{y=1}^{p-1}\left(\frac{y}{p}\right)\left(\frac{y^e-x^e g^s}{p}\right)^2.$$

If $s \neq 0$ the right member of the last equation reduces to $-\sum_{x,y=1}^{p-1}(xy/p)^{e+1}$, which is equal to 0 if e is even and is equal to $-(p-1)^2$ if e is odd. If $s=0$ we note that for a fixed value of x the value of the symbol $((y^e-x^e)/p)^2$ is equal to 0 if $y^e \equiv x^e \pmod p$ and is equal to 1 otherwise. In the proof of formula (2.4) we showed that when f is even the e roots of the congruence $y^e \equiv x^e \pmod p$ are given by $y_i \equiv r^i x \pmod p$, $i = 0, 1, \cdots, e-1$, where r belongs to the exponent $e \pmod p$ and $(r/p) = 1$. Hence

$$\begin{aligned}\sum_{x=1}^{p-1}\left(\frac{x}{p}\right)\sum_{y=1}^{p-1}\left(\frac{y}{p}\right)\left(\frac{y^e-x^e}{p}\right)^2 &= -\sum_{x=1}^{p-1}\left(\frac{x}{p}\right)\sum_{i=0}^{e-1}\left(\frac{r^i x}{p}\right) \\ &= -\sum_{x=1}^{p-1}\left(\frac{x}{p}\right)^2\sum_{i=0}^{e-1}\left(\frac{r}{p}\right)^i = -e(p-1).\end{aligned}$$

Combining our results we obtain Theorem 1.

The method of this section may be used to derive corresponding results for the sum $\psi_e(n)$. Thus we get for e even or odd

(4.4) $$\sum_{k=0}^{e-1}(-1)^k\psi_e(g^k) = -e,$$

which is analogous to (4.1). We may also prove

THEOREM 2. *Let s be a fixed integer such that $0 \leqq s \leqq e-1$. If e is even then*

$$(4.5)\qquad \sum_{k=0}^{e-1}\psi_e(g^k)\psi_e(g^{k+s})=\begin{cases} e^2p-2e(p-1) & (s=0),\\ -2e(p-1) & (s \text{ even}, s\neq 0),\\ 0 & (s \text{ odd}).\end{cases}$$

If e is odd then

$$(4.6)\qquad \sum_{k=0}^{e-1}\psi_e(g^k)\psi_e(g^{k+s})=\begin{cases} e^2p-e(p-1) & (s=0),\\ -e(p-1) & (s \text{ even}, s\neq 0),\\ e(p-1) & (s \text{ odd}).\end{cases}$$

5. Cyclotomy. The cyclotomic number (h,k) is the number of values of y, $1\leqq y\leqq p-2$, for which

$$(5.1)\qquad y\equiv g^{es+h},\qquad 1+y\equiv g^{et+k}\qquad (\mathrm{mod}\ p),$$

where the values of s and t are each selected from the integers $0,1,\cdots,f-1$.

Noting that $g^{ef}\equiv 1\ (\mathrm{mod}\ p)$ we may easily infer that

$$(5.2)\qquad (h,k)=(h+ae,k+be),$$

for any integers a and b. Furthermore it is not difficult to prove that

$$(5.3)\qquad (h,k)=\begin{cases} (e-h,k-h) & (f \text{ even or odd}),\\ (k,h) & (f \text{ even}),\\ (k+e/2,h+e/2) & (f \text{ odd}),\end{cases}$$

and

$$(5.4)\qquad \sum_{h=0}^{e-1}(h,k)=\begin{cases} f-1 & (k=0),\\ f & (1\leqq k\leqq e-1).\end{cases}$$

For a proof of (5.3) and (5.4) see, for example, Bachmann [1; Chapter 15].

We now consider the sum

$$(5.5)\qquad B(v,n)=\sum_{h=0}^{e-1}(h,v-nh).$$

This sum is equal to the number of values of y, $1\leqq y\leqq p-1$, for which $y^n(1+y)$ is congruent to a number of the form $x^eg^v(p\nmid x)$ with respect to the modulus p. Hence the number of solutions of the congruence

$$(5.6)\qquad y^{n+1}+y^n\equiv x^eg^v\ (\mathrm{mod}\ p)\qquad (v \text{ fixed};\ 0\leqq x,y\leqq p-1),$$

is equal to $2+eB(v,n)$. For a fixed integer a let $F_n(a)$ denote the number of values of y, $1\leqq y\leqq p-1$, for which $y^{n+1}+y^n\equiv a\ (\mathrm{mod}\ p)$. In terms of the functions $F_n(a)$ the number of solutions of the congruence (5.6) is equal to $\sum_{x=0}^{p-1}F_n(x^eg^v)$. Combining our results we conclude that

$$\sum_{x=0}^{p-1} F_n(x^e g^v) = 2 + eB(v, n). \tag{5.7}$$

The case $n = 1$ is an important special case. We may easily show that

$$F_1(x^e g^v) = 1 + \left(\frac{1 + 4g^v x^e}{p}\right),$$

so that (5.7) becomes

$$\sum_{x=0}^{p-1} \left(\frac{1 + 4g^v x^e}{p}\right) = -p + 2 + eB(v, 1).$$

The last equation leads at once to

$$eB(v, 1) = \begin{cases} p - 1 + \phi_e(4g^v) & (e \text{ odd}), \\ p - 1 + \psi_e(4g^v) & (e \text{ even}), \end{cases} \tag{5.8}$$

where $\phi_e(n)$ and $\psi_e(n)$ are defined by (2.1) and (3.1), respectively.

6. The case $e = 2$. As a first application let $e = 2$ and f be even, so that $p \equiv 1 \pmod 4$. In this case the second member of (4.3) reduces to the identity $p = a^2 + b^2$, where

$$a = \phi_2(1)/2, \qquad b = \phi_2(g)/2. \tag{6.1}$$

Next using (3.2), (3.9) with $e = 2$ and (3.10) with $e = 4$ we get

$$a \equiv -1 \pmod 4. \tag{6.2}$$

Formulas (6.1) and (6.2) constitute the theorem of Jacobsthal [5].

Applying (2.5) we obtain at once

$$2a \equiv -\binom{f}{f/2} \pmod p. \tag{6.3}$$

Formula (6.3) is a theorem of Gauss [3].

Again using (3.2) and (3.9) with $e = 2$ we get

$$a = (\psi_4(1) + 2)/2, \qquad b = \psi_4(g)/2. \tag{6.4}$$

Formula (6.4) is a theorem of Chowla [2; Theorem 1]. It may be remarked that (6.4) may also be deduced directly from (3.8), (4.4) and (4.5).

7. The case $e = 3$. If $p = 3f + 1$ the diophantine equation $4p = x^2 + 3y^2$ has three solutions in positive integers x and y. By (4.1) and (4.2) we get the equations

$$(7.1)\quad \phi_3(1)+\phi_3(g)+\phi_3(g^2)=-3,\quad \phi_3^2(1)+\phi_3^2(g)+\phi_3^2(g^2)=6p+3,$$

which may readily be transformed into the three identities

$$(7.2)\qquad 4p=(1+\phi_3(g^i))^2+3\left(\frac{\phi_3(g^{i+1})-\phi_3(g^{i+2})}{3}\right)^2 \qquad (i=0,1,2).$$

We now prove that the three solutions of $4p=x^2+3y^2$ given by (7.2) are distinct. For this purpose it suffices to show that

$$1+\phi_3(g^i)\neq\pm(1+\phi_3(g^j)) \qquad (i\neq j;\, i,j=0,1,2).$$

If $1+\phi_3(g^i)=1+\phi_3(g^j)$ then equations (7.1) lead to the absurd conclusion that $(1+\phi_3(g^i))^2=p$. If $1+\phi_3(g^i)=-(1+\phi_3(g^j))$ then the first equation in (7.1) implies that $1+\phi_3(g^k)=0$ for $k\neq i$, $k\neq j$. Hence by (7.2) we deduce $3\mid 4p$, which is impossible.

The diophantine equation $4p=c^2+3d^2$ has a unique solution with $c\equiv 1 \pmod 3$ and $d\equiv 0 \pmod 3$. This solution is given by

$$(7.3)\qquad c=1+\phi_3(4),\qquad d=(\phi_3(4g)-\phi_3(4g^2))/3.$$

To prove (7.3) we note that the congruence $c\equiv 1 \pmod 3$ follows at once from (2.4). Next using (5.2), (5.3), (5.5) and (5.8) we get

$$\phi_3(4g)=-p+1+9(0,1),\qquad \phi_3(4g^2)=-p+1+9(0,2),$$

so that $d\equiv 0 \pmod 3$. Formula (7.3) is the theorem of von Schrutka [7].

Applying (2.5) we obtain

$$(7.4)\qquad c\equiv -2^{2f}\binom{3f/2}{f/2}\equiv-\binom{2f}{f} \qquad (\operatorname{mod} p).$$

Formula (7.4) is due to Jacobi [4].

By (3.5), (3.7) and (3.9) with $e=3$ we get

$$2\phi_3(1)=\psi_6(1),\qquad \phi_3(g)-\phi_3(g^2)=\psi_6(g),$$

so that for $i=0$, (7.2) reduces to the identity $p=s^2+3t^2$, where

$$(7.5)\qquad s=(\psi_6(1)+2)/4,\qquad t=\psi_6(g)/6.$$

and the sign of s is determined by means of the congruence

$$(7.6)\qquad s\equiv -1 \qquad (\operatorname{mod} 3),$$

because of (3.10) with $e=6$. We have also by (2.5)

$$(7.7)\qquad 2s\equiv-\binom{3f/2}{f/2} \qquad (\operatorname{mod} p).$$

Formulas (7. 5) and (7. 6) comprise a theorem due to Chowla [2; Theorem 2]. We remark that (7. 5) may also be deduced directly from (3. 8), (4. 4) and (4. 5).

8. The case $e = 4$. In this section we consider the case $e = 4$ and f is even, so that $p \equiv 1 \pmod 8$. By (3. 7) and (3. 8) we get

$$\phi_4(g) = \phi_4(g^3),\ \phi_4(g^2) = 0; \qquad \phi_4^2(1) + \phi_4^2(g) + \phi_4^2(g^2) + \phi_4^2(g^3) = 16p.$$

Hence we have the identity $p = x^2 + 2y^2$, where

$$(8.1) \qquad x = \phi_4(1)/4, \qquad y = \phi_4(g)/4.$$

For another proof of (8. 1) see a recent paper [9; Theorem 2].

The sign of x in (8. 1) is determined by means of the congruence

$$(8.2) \qquad x \equiv (-1)^{f/2+1} \pmod 4.$$

To prove (8. 2) we return to (3. 11) and obtain the equation

$$\psi_4(1) = -p + 5 + 8A,$$

where $4A$ denotes the number of times that the symbol $((h^4+1)/p)$, $h = 1, 2, \cdots, p-1$, takes on the value 1. We next prove that A is an odd integer. Put $h \equiv g^k \pmod p$, $k = 0, 1, \cdots, p-2$. For $k = f/2, 3f/2, 5f/2, 7f/2$ we get $((g^{4k}+1)/p) = 0$. For $k = 0, f, 2f, 3f$ we get $((g^{4k}+1)/p) = (2/p) = 1$. For $k = 1, 2, \cdots, f/2-1, f/2+1, \cdots, f-1$ group the k's in pairs so that $f/2-j, f/2+j, j = 1, 2, \cdots, f/2-1$ form a pair. Since

$$(8.3) \quad (1 + g^{4(f/2-j)})(1 + g^{4(f/2+j)}) \equiv g^{4(f/2-j)}(1 + g^{4(f/2+j)})^2 \pmod p,$$

it follows that the two factors in the left member of (8. 3) have the same quadratic character. Therefore the number of values of k, $k = 0, 1, \cdots, f-1$, for which $((g^{4k}+1)/p)$ is equal to 1 is $2q+1$ for some integer q. Furthermore $g^{4(f/2\pm j)} \equiv g^{4(f/2+af\pm j)} \pmod p$ for any integer a. Hence $A = 2q+1$. This, in turn, leads to the congruence $\psi_4(1) \equiv (-1)^{f/2+1}4 \pmod{16}$. Using (3. 9) with $e = 4$ and (3. 10) with $e = 8$ we may now deduce (8. 2).

Applying (2. 5) we get easily

$$(8.4) \qquad 2x \equiv -\binom{2f}{f/2} \pmod p.$$

Formula (8. 4) is a theorem of Stern [8].

9. The case $e = 5$. For a prime $p = 5f + 1$, f is even. By (4.1) and (4.2) we have the identities

$$
(9.1) \qquad \begin{aligned}
&\sum_{k=0}^{4} \phi_5(g^k) = -5, \qquad \sum_{k=0}^{4} \phi_5{}^2(g^k) = 20p + 5, \\
&\sum_{k=0}^{4} \phi_5(g^k)\phi_5(g^{k+1}) = \sum_{k=0}^{4} \phi_5(g^k)\phi_5(g^{k+2}) = -5(p-1).
\end{aligned}
$$

After some manipulation we may transform the identities in (9.1) into the identities

$$
(9.2) \qquad \begin{aligned}
&16p = A_k{}^2 + 5B_k{}^2 + 10C_k{}^2 + 10D_k{}^2 \qquad (k = 0, 1, 2, 3, 4), \\
&A_kB_k = C_k{}^2 - 4C_kD_k - D_k{}^2,
\end{aligned}
$$

where

$$
(9.3) \qquad \begin{aligned}
&A_k = 1 + \phi_5(g^k), \ 5B_k = \phi_5(g^{k+1}) - \phi_5(g^{k+2}) - \phi_5(g^{k+3}) + \phi_5({}^{k+4}), \\
&5C_k = \phi_5(g^{k+2}) - \phi_5(g^{k+3}), \ 5D_k = \phi_5(g^{k+1}) - \phi_5(g^{k+4}).
\end{aligned}
$$

Another pair of identities is

$$
(9.4) \qquad 16p = x^2 + 50u^2 + 50v^2 + 125w^2, \qquad xw = v^2 - 4uv - u^2,
$$

where

$$
(9.5) \qquad \begin{aligned}
x &= 1 + \phi_5(4), \\
25\mathrm{u} &= \phi_5(4g) + 2\phi_5(4g^2) - 2\phi_5(4g^3) - \phi_5(4g^4), \\
25v &= 2\phi_5(4g) - \phi_5(4g^2) + \phi_5(4g^3) - 2\phi_5(4g^4), \\
25w &= \phi_5(4g) - \phi_5(4g^2) - \phi_5(4g^3) + \phi_5(4g^4),
\end{aligned}
$$

and the sign of x is determined by means of the congruence $x \equiv 1 \pmod 5$ in view of (2.4).

The formula for x is due to Emma Lehmer [6]. To prove (9.4) select k in (9.3) so that $g^k \equiv 4 \pmod p$. Making use of the formulas in § 5 we may express u, v and w defined in (9.5) in terms of the cyclotomic numbers (h, k). Dropping subscripts in (9.3) we arrive at the equations

$$
\begin{aligned}
25u &= 10C + 5D = 25[(0, 2) - (0, 3)], \\
25v &= -5C + 10D = 25[(0, 1) - (0, 4)], \\
25w &= 5B = 25[(1, 3) - (1, 2)].
\end{aligned}
$$

Hence $A = x$, $B = 5w$, $C = 2u - v$, $D = u + 2v$ and (9.2) reduces to (9.4).

UNIVERSITY OF SOUTHERN CALIFORNIA.

REFERENCES.

[1] P. Bachmann, *Die Lehre von der Kreistheilung*, Leipzig-Berlin, 1921.

[2] S. Chowla, "The last entry in Gauss's diary," *Proceedings of the National Academy of Sciences*, vol. 35 (1949), pp. 244-246.

[3] K. F. Gauss, "Theorie residuorum biquadraticorum," *Werke*, vol. 2, p. 90.

[4] K. Jacobi, "De residuis cubicis commentatio numersoa," *Journal für die reine und angewandte Mathematik*, vol. 2 (1827), pp. 66-69.

[5] E. Jacobsthal, "Über die Darstellung der Primzahlen der Form $4n+1$ als Summe zweier Quadrate," *Journal für die reine und angewandte Mathematik*, vol. 132 (1907), pp. 238-245.

[6] E. Lehmer, "On the quintic character of 2," *Bulletin of the American Mathematical Society*, vol. 55 (1949), Abstract No. 72, p. 62.

[7] L. von Schrutka, "Eine Beweis für die Zerlegbarkeit der Primzahlen von der Form $6n+1$ in ein einfaches und ein dreifaches Quadrat," *Journal für die reine und angewandte Mathematik*, vol 140 (1911), pp. 252-265.

[8] M. Stern, "Eine Bemerkung zur Zahlentheorie," *Journal für die reine und angewandte Mathematik*, vol. 32 (1846), pp. 89-90.

[9] A. L. Whiteman, "Theorems analogous to Jacobsthal's theorem," *Duke Mathematical Journal*, vol. 16 (1949), pp. 619-626.

INFINITELY NEAR POINTS ON ALGEBRAIC SURFACES.*

By GINO TURRIN.[1]

1. Infinitely near points on plane algebraic curves were introduced by Max Noether [1] in the process of reduction of singularities. Noether's classical theorem is: a sequence of singular infinitely near points on a plane algebraic curve is necessarily finite. In the case of space curves and surfaces the first definitions and results were established by C. Segre [2] using quadratic transformations of the ambient space. Noether's result, though it remains valid for space curves, is no longer true for surfaces because of the appearance of singular curves.

As a general rule, and even in the simplest case of plane algebraic curves, the analysis of infinitely near points is intricate and all proofs have to be carried out by handling a great number of details.[2] In the case of algebraic surfaces B. Levi [3] proved the following theorem: *if $P, P_1, P_2, \cdots$ is an infinite sequence of infinitely near points on an algebraic surface and all of the same multiplicity $\nu > 1$, then for any given p there exists $q > p$ such that the point P_q lies on the transform of a ν-fold curve passing through the point P_{q-1} immediately preceding P_q.* It is the purpose of this note to give a purely algebraic proof of Levi's theorem for arbitrary ground fields of characteristic zero. Our proof makes use of related results proved in the fundamental paper of Zariski [7] and it goes further by showing the existence of an index p such that for *any* $q > p$ the point P_q lies on a ν-fold curve.

2. A trivial case of an infinite sequence of infinitely near points which all have the same multiplicity $\nu > 1$ is one in which there exists an index p such that the point P_p lies on a ν-fold curve Γ_p and all the points sucessive to P_p lie on the corresponding transforms of the curve Γ_p. A sequence of this sort will be called a *trivial sequence.*

We shall need the following remark: *Let P be a point of a surface F and let T be a quadratic transformation of center P which sends F into a surface F_1 and the fundamental point P into the curve $\Delta = T[P]$ of F.*

* Received October 5, 1950.

[1] Fellow of the Rockefeller Foundation.

[2] See for example [5] or [6], chap. 1.

If Γ_1 is an irreducible curve of F_1 not contained as component in Δ and such that $\Gamma_1 \cap \Delta \neq 0$, then Γ_1 is the transform $T[\Gamma]$ of an irreducible curve Γ of F passing through P and Γ_1 has for F_1 the same multiplicity as Γ has for F. This is immediate: since P is the only fundamental point of the transformation T and since Γ_1 is not a component of $\Delta - T[P]$ it follows that Γ_1 is the transform $T[\Gamma]$ of an irreducible curve Γ of the surface F. Since under the inverse transformation T^{-1} to any point of Δ corresponds the point P and since $\Gamma_1 \cap \Delta \neq 0$ we conclude that Γ contains the point P. Finally, the respective multiplicities of Γ and Γ_1 are the same since the quadratic transformation T is regular at a general point of the curve Γ.

3. Our result is as follows:

THEOREM. *Let $P, P_1, P_2, \cdots$ be an infinite sequence of infinitely near points all of the same multiplicity $\nu > 1$ on an irreducible algebraic surface F defined over an arbitrary ground field of characteristic zero. Then there exists an index p such that for any $q > p$ the point P_q lies on at least one irreducible curve Γ_q which is ν-fold for the surface F_q containing P_q. The curve Γ_q either* a) *coincides with the transform $T_{q-1}[P_{q-1}]$ of the point P_{q-1} under the quadratic transformation $T_{q-1}\colon F_{q-1} \to F_q$, or* b) *is the transform $T_{q-1}[\Gamma_{q-1}]$ of some curve Γ_{q-1} passing through P_{q-1} and ν-fold for the surface F_{q-1}. Case* b) *must occur infinitely many times.*

Proof. A given sequence $P, P_1, P_2, \cdots$ of points infinitely near to P on the surface F determines at least one valuation v of the field of algebraic functions of which the birationally equivalent surfaces $F, F_1, F_2, \cdots$ are distinct projective models. If the given infinite sequence $P, P_1, P_2, \cdots$ is a trivial sequence the Theorem is obviously true. Therefore we assume that *the sequence $P, P_1, P_2, \cdots$ is not trivial and all the points have the same multiplicity $\nu > 1$.* Zariski [7] has proved that under these conditions the valuation v must be 0-dimensional non-discrete of rank one and that moreover, after a finite number of quadratic transformations, the element ω_i which defines the surface F_i in the quotient ring $\mathfrak{J}_i$ of the point P_i with respect to the 3-dimensional ambient space of F_i can be written in the following form: (we may assume that $P_i = P$ and $F_i = F$ are the initial point and surface respectively)

$$\omega = \epsilon_0 z^\nu + \sum_{j=1}^{\nu} \epsilon_j z^{\nu-j} x^{m_{\alpha j}} y^{n_{\alpha j}} \tag{1}$$

where the ϵ_j, $0 \leqq j \leqq \nu$, are either zero or units of the quotient ring $\mathfrak{J}$ (but $\epsilon_0 \neq 0$) and where z, x, y is a set of uniformizing parameters of the point P

at $\mathfrak{J}$. Since the point $P=(0,0,0)$ is of multiplicity ν for the surface F the exponents which appear in (1) must satisfy the inequalities $m_{0j}+n_{0j}\geqq j$, for all $j\neq 0$ such that $\epsilon_j\neq 0$. Furthermore, *loc. cit.* Lemma 11.4, the leading form of the element ω is the ν-th power of a linear form. Taking into account this fact and the expression (1) we shall obtain the stronger inequalities

$$m_{0j}+n_{0j}>j \tag{2}$$

for all $j\neq 0$ such that $\epsilon_j\neq 0$. For let $\mathfrak{L}=(a_1\bar{z}+a_2\bar{y}+a_3\bar{x})^\nu$ be the leading form of ω where a_1, a_2, a_3 are elements of the residue field $\mathfrak{K}$ of P, $\bar{z},\bar{y},\bar{x}$ are algebraically independent transcendentals over $\mathfrak{K}$, and where $a_1\neq 0$ since $\epsilon_0\neq 0$. If, say, $a_2\neq 0$ then necessarily $a_3=0$ since otherwise the expression of $\mathfrak{L}$ as a sum of monomials would contain $(\nu+1)(\nu+2)/2$ $(>\nu+1)$ terms contradicting the expression (1) for ω. Thus we may assume $a_1, a_2\neq 0$ and $a_3=0$. Now the expression for $\mathfrak{L}$ as a sum of monomials has exactly $\nu+1$ terms. This implies in (1) that all ϵ_j are different from zero and also that the total degree in the uniformizing parameters of each monomial of ω is exactly ν. Since $\bar{x}$ does not occur in $\mathfrak{L}$, the expression (1) for ω takes the form $\omega=\epsilon_0 z^\nu+\epsilon_1 z^{\nu-1}y+\cdots+\epsilon_\nu y^\nu$. Let α_1 and α_2 be elements of $\mathfrak{J}$ whose residues coincide with a_1 and a_2 respectively. Then the leading form of the element $\omega'=(\alpha_1 z+\alpha_2 y)^\nu$ coincides with $\mathfrak{L}$ and since the quotient ring $\mathfrak{J}$ of the point P with respect to the ambient 3-dimensional space is a regular local ring we conclude that $\omega=\omega'$. This contradicts the fact that F is an irreducible surface. Hence necessarily also $a_2=0$, i.e., $\mathfrak{L}=(a_1\bar{z})^\nu$ which, by definition of the leading form, shows the validity of the inequalities (2) for the expression (1).

We point out that the expression (1) together with conditions (2) is permanent in the sense that all the elements ω_i which succeed ω when the successive quadratic transformations are performed have expressions similar to (1) in which the corresponding conditions analogous to (2) hold. In order to write the local equations of the quadratic transformation T which sends the surface F defined in $\mathfrak{J}$ by ω into the surface F_1 defined in $\mathfrak{J}_1$ by ω_1, and thus to obtain for ω_1 the expression similar to (1) in terms of the uniformizing parameters z_1, y_1, x_1 of the point P_1, we must distinguish three cases according to the ratio of the values assigned to y and x by the fixed valuation v:

Case 1. $v(x)<v(y)$. The local equations of T are:

$$z_1=z/x;\qquad y_1=y/x,\qquad x_1=x; \tag{4.1}$$

consequently

$$\omega_1 = \omega/x^\nu = \epsilon_0 z_1^\nu + \sum_{j=1}^{\nu} \epsilon_j z_1^{\nu-j} x_1^{m_{1j}} y_1^{n_{1j}} \tag{1.1}$$

where

$$m_{1j} = m_{0j} + n_{0j} - j, \qquad n_{1j} = n_{0j} \tag{3.1}$$

and where the conditions similar to (2) are

$$m_{1j} + n_{1j} > j. \tag{2.1}$$

Case 2. $v(x) = v(y)$. Since the valuation v has center in all the quotient rings $\mathfrak{J}_i$, in particular in $\mathfrak{J}_1$, the element y/x is a unit in $\mathfrak{J}_1$ and we have:

$$z_1 = z/x, \qquad y_1 = f(y/x), \qquad x_1 = x \tag{4.2}$$

where $f(y/x)$ is a suitable polynomial in y/x with coefficients in $\mathfrak{J}$.[8]

$$\omega_1 = \omega/x^\nu = \epsilon_0 z_1^\nu + \sum_{j=1}^{\nu} \bar{\epsilon}_j z_1^{\nu-j} x_1^{m_{1j}} \tag{1.2}$$

where $\bar{\epsilon}_j = \epsilon_j \cdot (y/x)^{n_{0j}}$ is clearly a unit in $\mathfrak{J}_1$ (if $\neq 0$) and where

$$m_{1j} = m_{0j} + n_{0j} - j, \qquad (n_{1j} = 0) \tag{3.2}$$

and

$$m_{1j} > j. \tag{2.2}$$

Case 3. $v(x) > v(y)$. In this case we have:

$$z_1 = z/y, \qquad y_1 = y, \qquad x_1 = x/y, \tag{4.3}$$

$$\omega_1 = \omega/y^\nu = \epsilon_0 z_1^\nu + \sum_{j=1}^{\nu} \epsilon_j z_1^{\nu-j} x_1^{m_{1j}} y_1^{n_{1j}}, \tag{1.3}$$

$$m_{1j} = m_{0j}, \qquad n_{1j} = n_{0j} + m_{0j} - j, \tag{3.3}$$

and as usual

$$m_{1j} + n_{1j} > j. \tag{2.3}$$

It follows from (1), (2) and (4. s) ($s = 1, 2, 3$) that the curve $\Delta = T[P]$ which corresponds on the surface F_1 to the fundamental point $P = (0, 0, 0)$ is irreducible and is given in the quotient ring $\mathfrak{J}_1$ of P_1 by the ideal (z_1, x_1) in cases 1 and 2, and by the ideal (z_1, y_1) in case 3. We shall also make use of the fact that in case 3 the curve given in $\mathfrak{J}_1$ by the

[8] If the ground field is algebraically closed $f(y/x) = y/x - c$, where c is the v-residue of y/x.

ideal (z_1, x_1) is merely the transform of the curve which passes through the point $P = (0, 0, 0)$ and is given in $\mathfrak{J}$ by the ideal (z, x).

With reference to the first part of the Theorem we observe first of all that *given any index p there exists $q > p$ such that the point P_q is not isolated.*[4] In fact, our hypothesis that $P, P_1, P_2, \cdots$ is a non-trivial sequence of points all of the same multiplicity $\nu > 1$ implies that the local uniformization of the 0-dimensional valuation v ([7] Theorem 5) cannot be performed by using quadratic transformations only. Since the process of uniformization of v is carried out by a finite number of quadratic and monoidal transformations and since monoidal transformations come into play only when some intermediate point P_i is not isolated, our assertion follows. Therefore we may asume that in the surface F given by (1) there is a ν-fold curve passing through the point $P = (0, 0, 0)$. Moreover, *we may assume that the ν-fold curve is given in the quotient ring $\mathfrak{J}$ of P by the ideal (z, x).* For if no point after P is isolated the first part of the Theorem is already proved. Otherwise since we have just shown that in our given infinite sequence of points there exists an infinite number of points which are not isolated, we may suppose for a moment that the point P is isolated and that P_1 is not isolated. By the remark of § 2 the ν-fold curve through P_1 must coincide with the transform $\Delta = T[P]$ of the point P. Now, we have seen above that Δ is given in the quotient ring $\mathfrak{J}_1$ of P_1 either by the ideal (z_1, x_1) (cases 1 and 2) or by the ideal (z_1, y_1) (case 3). Hence, by setting again $P_1 = P$ and interchanging, if necessary, the notation for the two uniformizing parameters x and y, we prove the assertion. Therefore we assume that in the surface F given by (1) the curve defined in the quotient ring $\mathfrak{J}$ by the ideal (z, x) is ν-fold for F. This is equivalent to saying that in (1) the following additional conditions to (2) hold:

$$m_{0j} \geqq j \tag{2'}$$

for all $j \neq 0$ such that $\epsilon_j \neq 0$. Once all the above assumptions are made the first part of the Theorem would be proved when it is shown that *no matter which case* (1, 2 or 3) *occurs, then in the corresponding expression for ω_1 we necessarily have $m_{1j} \geqq j$ for all $j \neq 0$ such that $\epsilon_j \neq 0$.* This is given by (2. 2) in case 2 and is an obvious consequence of (2′) and the first part of (3. 3) in case 3. To prove the assertion for case 1 we need the following general remark applicable to the expression (1): *for no $j \neq 0$ such that $\epsilon_j \neq 0$ can it happen that*

$$\text{(A)} \quad m_{0j} < j \qquad\qquad \text{(B)} \quad n_{0j} < j$$

[4] A ν-fold point P is isolated if no curve of the surface passing through P is ν-fold.

simultaneously. For whatever case 1, 2 or 3 occurs as the next step, it is clear from (A), (B) and (3. s) ($s=1,2,3$) that

$$m_{0j}+n_{0j}>m_{1j}+n_{1j}$$

and that again, using once more (A), (B) and (3. s), we have simultaneously

$$(\mathrm{A}_1)\quad m_{1j}<j \qquad\qquad (\mathrm{B}_1)\quad n_{1j}<j.$$

Thus we should get an infinite strictly decreasing sequence of inequalities

$$m_{0j}+n_{0j}>m_{1j}+n_{1j}>m_{2j}+n_{2j}>\cdots$$

contradicting the generally valid inequalities similar to (2). Now, if in (3. 1) we had, for some $j\neq 0$ for which $\epsilon_j\neq 0$, $m_{1j}<j$ it would follow from (2′) and (3. 1) also that $n_{1j}<j$. This we have seen to be impossible and hence also in case 1 we must necessarily have $m_{1j}\geqq j$ for all $j\neq 0$ such that $\epsilon_j\neq 0$. This completes the proof of the first part of the Theorem. The second part of the Theorem is evident from the remark of § 2 and together with the first part represents our improvement of Levi's result in the sense that there exists a certain point in the sequence such that *all* its successors are non-isolated and the ν-fold curves which appear through the points are of a certain well determined type.

To prove the last part of the Theorem we shall now show that, *given an index p there exists $q>p$ such that the point P_q lies on the transform of a ν-fold curve passing through P_{q-1}.* Returning to the expression (1) and under the assumption (2′), i. e., the condition that the curve defined by the ideal (z,x) is ν-fold for the surface F, we shall consider the order of consecutive appearance of the three cases 1, 2 and 3. We have just shown in the proof of the first part of the Theorem that, independently of the different cases, the curve of the surface F_1 given by the ideal (z_1,x_1) is ν-fold. If case 3 occurs infinitely many times our assertion is proved since the curve given by the ideal (z_1,x_1) is in that case the transform of the ν-fold curve given by the ideal (z,x). Now, case 2 cannot take place infinitely many times *consecutively* since by (3. 2) the exponents of the successive parameters x_i decrease strictly. It is also clear from (1. 2) that case 2 or case 3 are the only ones which can occur immediately after case 2. Therefore it remains only to show that case 1 cannot appear infinitely many times consecutively. This is seen as follows: if case 1 occurs i times consecutively (see (4. 1)) the values of the parameters are related in the following form; $0<v(x_i)=v(x)$ and $0<v(y_i)=v(y)-iv(x)$. Since v is of rank one $v(x)$ and $v(y)$ may be supposed to be real numbers. Hence $i<v(y)/v(x)$. This completes the proof of the Theorem.

4. It is well known by construction of examples that there exist infinite sequences of infinitely near points on an algebraic surface all of the same multipilicty which are not trivial sequences; see [4] or [6], p. 15. Recently Derwidué [5] has made a new attempt at a purely geometric proof of the theorem of reduction of singularities of an algebraic surface by using Cremona transformations of the ambient 3-space. The proof makes use of a statement concerning the behaviour of the base points of the polar curves and it is asserted that from that statement the following result attributed to B. Levi follows as a simple corollary: an infinite sequence of infinitely near singular points on an algebraic surface is necessarily a trivial sequence. As we have just shown above this assertion is neither true nor does it represent B. Levi's theorem. The procedure follows the classical approach of eliminating the isolated singularities and further reducing the multiple curves but fails to show the true fact that the alternate process of: a) reduction of singular curves, b) elimination of the isolated singularities introduced by a), a_1) reduction of the singular curves introduced by b), etc. is necessarily finite.

HARVARD UNIVERSITY.

REFERENCES.

[1] M. Noether, "Ueber die singulären Werthsysteme einer algebraischen Function und die singulären Punkte einer algebraischen Curve," *Mathematische Annalen*, vol. 9 (1876), pp. 166-182.

[2] C. Segre, "Sulla scomposizione dei punti singolari delle superficie algebriche," *Annali di Matematica*, serie II, vol. 25 (1897), pp. 1-53.

[3] B. Levi, "Sulla riduzione delle singolarita puntuali delle superficie algebriche dello spazio ordinario per trasformazioni quadratiche," *Annali di Matematica*, serie II, vol. 26 (1897), pp. 219-253.

[4] B. Levi, "Intorno alla composizione dei punti generici delle linee singolari delle superficie algebriche," *Annali di Matematica*, serie III, vol. 2 (1899), pp. 127-138.

[5] Enriques-Chisini, *Lezioni sulla teoria geometrica delle equazioni e delle funzioni algebriche*, vol. 2, Zanichelli, 1915-1924.

[6] O. Zariski, "Algebraic surfaces," *Ergebnisse der Mathematik*, vol. 3 (1935), pp. 1-198.

[7] O. Zariski, "Reduction of the singularities of algebraic three dimensional varieties," *Annals of Mathematics*, vol. 45 (1944), pp. 472-542.

---

[5] L. Derwidué, "Résolution des singularités d'une surface algébrique au moyen de transformations crémoniennes," *Bull. Soc. Roy. Sci. Liège*, vol. 16 (1947), pp. 275-289.

SOME EXAMPLES IN THE THEORY OF SINGULAR BOUNDARY VALUE PROBLEMS.*

By PHILIP HARTMAN.**

1. The theorems. Let $q(t)$, where $0 \leqq t < \infty$, be a real-valued, continuous function and λ be a real parameter. Let $N(T, \lambda)$ denote the number of zeros on $0 < t < T$ of a solution $x = x(t) = x(t, \lambda) \not\equiv 0$ of the differential equation

$$x'' + (q + \lambda)x = 0. \tag{1}$$

Thus, up to an additive correction -1, 0 or $+1$ (depending on T), the number $N(T, \lambda)$ is independent of the particular solution $x = x(t)$ of (1) determining it. For example, if $q(t) \equiv 0$, then, as $T \rightarrow \infty$,

$$N(T, \lambda) = O(1) \text{ if } \lambda \leqq 0; \qquad N(T, \lambda) = \lambda^{\frac{1}{2}}T/\pi + O(1) \text{ if } \lambda > 0. \tag{2}$$

The asymptotic behavior, as $T \rightarrow \infty$, of $N(T, \lambda)$ can depend in a very complicated manner on λ, even for simple functions $q(t)$. To illustrate this, let $q(t)$ be a periodic function, say $q(t) = \cos t$; the solution of the problem of the asymptotic behavior of $N(T, \lambda)$ furnishes the solution of the problem of the determination of the λ-values for which (1) has periodic or half-periodic solutions. Thus, in general, one cannot expect a solution as simple as (2). Nevertheless, it seems surprising that the situation can be as pathological as indicated by the following theorem:

(*) *Let $\psi = \psi(T)$ be a positive, continuous, non-decreasing function for $0 < T < \infty$. Then there exist real-valued, continuous functions $q = q(t)$ for $0 \leqq t < \infty$ realizing each of the following situations, as $T \rightarrow \infty$:*

$$N(T, \lambda = O(1) \text{ if } \lambda < 0; \qquad N(T, \lambda) = \psi(T) + O(1) \text{ if } \lambda > 0, \tag{3}$$

$$N(T, \lambda) = \psi(T) + O(1) \text{ for all } \lambda, \tag{4}$$

$$N(T, \lambda) = \psi(T) + O(1) \text{ if } \lambda < 0; \qquad N(T, \lambda) = 2\psi(T) + O(1) \text{ if } \lambda > 0. \tag{5}$$

It is curious that ψ can tend slowly or rapidly to ∞ with T, while the

* Received January 23, 1951.

** John Simon Guggenheim Memorial Foundation Fellow, on leave of absence from The Johns Hopkins University.

right-hand sides of the asymptotic formulae in (3), (4), (5) are essentially independent of λ. The contrast furnished by (2) is indeed great.

Remark on $\lambda = 0$. In the examples, constructed below for the cases (3) and (5) of (*), the number $\lambda = 0$ satisfies the second relation in (3) and (5), respectively; that is, as $T \to \infty$,

(3 bis) $$N(T, 0) = \psi(T) + O(1),$$

(5 bis) $$N(T, 0) = 2\psi(T) + O(1).$$

But it turns out that any situation consistent with Sturm's comparison theorem can be realized. Thus, if $\phi(T)$ is any positive, continuous, non-decreasing function for $0 < T < \infty$ satisfying $\phi(T_2) - \phi(T_1) \leqq \psi(T_2) - \psi(T_1)$ for $0 < T_1 < T_2 < \infty$, then examples $q(t)$ realizing (3) or (5), respectively, can be chosen so that

(3_0) $$N(T, 0) = \phi(T) + O(T)$$

or

(5_0) $$N(T, 0) = \phi(T) + \psi(T) + O(T).$$

A modification of the proof of (*) yields:

(**) *Let S be a closed set on the λ-axis. There exist real-valued, continuous functions $q = q(t)$ on $0 \leqq t < \infty$ such that, for every pair of numbers λ, μ, the difference $N(T, \mu) - N(T, \lambda)$ is unbounded or bounded, as $T \to \infty$, according as the open interval (λ, μ) contains points of S or the closed interval $[\lambda, \mu]$ does not contain points of S.*

Remark. If S is unbounded from below, then $N(T, \lambda) \to \infty$, as $T \to \infty$, for every λ. If S is bounded from below, then examples $q(t)$, proving (**), can be chosen so that each of the following alternatives is realized: as $T \to \infty$, $N(T, \lambda) \to \infty$ for every λ or $N(T, \lambda) = O(1)$ if λ is less than every number in S.

The proofs of (*) and (**) will make it clear that the "pathology" in the examples is not associated with the local smoothness of $q(t)$. In fact, (*) and (**) remain true if the phrase "continuous functions $q(t)$" is replaced by "functions $q(t)$ of class C^∞."

The theorem (*) furnishes the solution to the problem, suggested to me by Professor Wintner, of characterizing the monotone functions $\psi(T)$ corresponding to which there exist differential equations (1) satisfying, as $T \to \infty$, $N(T, \lambda) = O(1)$ if $\lambda < 0$ and $N(T, \lambda) = o(\psi(T))$ if $\lambda > 0$; cf. (3). The

theorems (*) and (**) also make it possible to answer some questions, raised by him, concerning the possible structure of the spectra of differential operators (e. g., the question whether the tacit assumption, made by various writers, that the spectrum in the completely continuous case clusters only at ∞, and not at $-\infty$ as well, is not a mistaken one). The answers to these questions are contained in some of the corollaries of the following section.

2. Consequences for the spectral theory of (1). The differential equation (1), for a fixed λ, is said to be *oscillatory* or *non-oscillatory* according as every solution of (1) does or does not have an infinity of zeros. The differential equation (1) is said to be of *limit-circle* or *limit-point* type according as (1) does or does not have two linearly independent solutions of class $L^2(0,\infty)$; the type of (1) is independent of λ; [9], p. 238. In the limit-point case, (1) and a linear homogeneous boundary condition at $t=0$,

$$x(0)\cos\alpha - x'(0)\sin\alpha = 0, \qquad (0 \leqq \alpha < \pi), \tag{6}$$

determine a self-adjoint boundary value problem in $L^2(0,\infty)$; [9]. Let S_α denote the spectrum of this problem and let S' denote the set of (finite) cluster points of S_α. The set S', "the essential spectrum," is independent of α; [9], p. 251. The characterization [2] of S_α and S' in terms of $N(T,\lambda)$, together with (*), (**), gives some curious consequences for the theory of the spectra of differential operators.

No examples are known for which (1) is of limit-point type and S_α contains a point spectrum which clusters at $-\infty$. However, (*) implies the following:

COROLLARY 1. *There exist real-valued, continuous functions $q=q(t)$ on $0\leqq t<\infty$ such that* (1) *is of limit-point type and S_α (for every α) is a pure point spectrum clustering at, and only at, both ∞ and $-\infty$.*

The spectrum S_α has no finite cluster if and only if the Green kernel, belonging to S_α and a λ (real or non-real) in the complement of S_α, is completely continuous. The known examples of completely continuous Green kernels are the cases where (1) is non-oscillatory for every λ; correspondingly S_α clusters only at ∞. The Green kernels cannot have the property of complete continuity if (1) is oscillatory for some λ and non-oscillatory for some λ; [3]. But examples, proving Corollary 1, show that certain Green kernels can have this property when (1) is oscillatory for every λ (and S_α clusters at both $-\infty$ and ∞).

COROLLARY 2. *There exist real-valued, continuous functions* $q = q(t)$ *on* $0 \leqq t < \infty$ *such that* (1) *is of limit-point type, is oscillatory for every* λ *and possesses a non-trivial* ($\not\equiv 0$) *solution of class* $L^2(0, \infty)$ *for every* λ. *Furthermore, the "Green kernels" are completely continuous.*

It is known that S_α always cluster at ∞; [5], p. 310. It is also known [11] that if S_α contains a (non-vacuous) continuous spectrum, then the latter is unbounded. However, it is not known whether or not the essential spectrum S' is necessarily unbounded when it is non-vacuous. Nor is it known whether or not S' can contain an isolated point. The latter two questions are answered by

COROLLARY 3. *There exist real-valued, continuous functions* $q = q(t)$ *on* $0 \leqq t < \infty$ *such that* (1) *is of limit-point type and* S_α (*for every* α) *is a pure point spectrum clustering at, and only at,* 0 *and* ∞ (*or at, and only at,* 0 *and both* $-\infty$ *and* ∞).

If (1) is oscillatory for some λ and non-oscillatory for some other λ, then (1) is of limit-point type ([1]) and the least cluster point of S_α is the greatest lower bound, λ_0, of those λ for which (1) is oscillatory; [3]. The spectrum S_α clusters at $\lambda = \lambda_0$ from the left if and only if (1) is oscillatory for $\lambda = \lambda_0$; [1], p. 698. In the known examples with $-\infty < \lambda_0 < \infty$, the spectrum S_α always clusters at $\lambda = \lambda_0$ from the right. But it can be shown that this need not be the case in general.

COROLLARY 4. *There exist real-valued, continuous functions* $q = q(t)$ *on* $0 \leqq t < \infty$ *such that* (1) *is of limit-point type, the essential spectrum* S' *consists of the single point* $\lambda = 0$, *but no interval* $(0, \lambda)$, *where* $\lambda > 0$, *contains an infinite subset of* S_α (*for any fixed* α).

Another consequence of (*) is the negative assertion in

COROLLARY 5. *A necessary, but not sufficient, condition that* (1) *be of limit-circle type is that* (1) *be oscillatory and that*

$$N(T, \lambda_2) - N(T, \lambda_1) = O(1), \text{ as } T \to \infty, \tag{7}$$

for every pair of λ*-values* λ_1, λ_2.

In view of the last corollary and the characterization of essential spectra given in [2], the theorem (**) implies:

COROLLARY 6. *Let* S *be a closed set on the* λ*-axis. There exist real-valued, continuous functions* $q = q(t)$ *on* $0 \leqq t < \infty$ *such that* (1) *is of limit-point type and the essential spectrum* S' *of* (1) *is the given set* S.

For example, if S is a perfect, nowhere dense set on the interval $0 \leqq \lambda \leqq 1$, then there exist differential equations (1) of limit-point type for which the set of cluster points of S_α, for every fixed α, is S and the point ∞ or S and the points ∞ and $-\infty$. (Each of the alternatives can be realized by virtue of the Remark following (**).)

In the case just mentioned, or whenever S is bounded, there cannot exist any continuous spectrum, that is, S_α is a pure point spectrum; [11]. Thus, if S is chosen to be the entire interval $0 \leqq \lambda \leqq 1$, one obtains

COROLLARY 7. *There exist real-valued, continuous functions $q = q(t)$ on $0 \leqq t < \infty$ such that* (1) *is of limit-point type and, for every fixed α, the spectrum S_α is a pure point-spectrum which is dense on $0 \leqq \lambda \leqq 1$ and has ∞ (or ∞ and $-\infty$) as its only cluster points not on the interval $0 \leqq \lambda \leqq 1$.*

It is curious to compare this with the result of [5], p. 660.

3. Proof of Corollaries 1 and 2. Choose $\psi(T) = T$ in (*) and let $q = q(t)$ be such that (4) holds. Then it follows that $N(T, \lambda) \sim T = O(T^2)$, as $T \to \infty$, for any λ. Hence, by [6], (1) is of limit-point type. It also follows from (4) that, if $-\infty < \lambda_1 < \lambda_2 < \infty$, then $N(T, \lambda_2) - N(T, \lambda_1) = O(1)$, as $T \to \infty$. Consequently, the characterization of S_α given in [2] implies that S_α has no finite cluster point, that is, that S' is empty. Hence, S_α is a pure point spectrum. But S_α always cluster at ∞; [5], p. 310. Also, S_α clusters at $-\infty$ whenever (1) is oscillatory for every λ; [5], pp. 313-314. This proves Corollary 1.

Corollary 2 is an immediate consequence of Corollary 1. For if λ is not in S', then (1) possesses a non-trivial ($\not\equiv 0$) solution of class $L^2(0, \infty)$; [4]. The complete continuity of the "Green kernels" follows from the remarks preceding the statement of Corollary 2.

4. Proof of Corollary 3. Choose $\psi(T)$ so that $\psi(\infty) = \infty$, and $q(t)$ so that (3) holds. Then (1) is of limit-point type by virtue of the first part of (3); [1]. That $\lambda = 0$ is a cluster point (and the least cluster point $\geqq -\infty$) of S_α follows from [3]. Also $N(T, \lambda_2) - N(T, \lambda_1)$ is bounded or unbounded, as $T \to \infty$, according as λ_1, λ_2 are or are not of the same sign. Consequently, $\lambda = 0$ is the only finite cluster point of S_α, by [2]. Since S_α always clusters at ∞, the proof of the first part of Corollary 3 is complete.

In order to prove the last (parenthetical) part of the Corollary, let $\psi(T) = T$ and let $q(t)$ be such that (5) holds. Then, as in the last section, (1) is of limit-point type and S_α clusters at $\pm\infty$. As in the last paragraph,

$\lambda = 0$ is the only finite cluster point of S_α. This completes the proof of Corollary 3.

5. Proof of Corollary 4. This corollary is a consequence of the proof of the first part of Corollary 3 and the Remark on $\lambda = 0$ in § 1. Let (3 bis) hold, as well (3). Then it follows that $N(T, \lambda) - N(T, 0) = O(1)$, as $T \to \infty$, if $\lambda > 0$. Thus, by [2], the interval $(0, \lambda)$ contains at most a finite number of points of S_α (for any fixed α).

6. Proof of Corollary 5. That a necessary condition that (1) may be of limit-circle type is that (1) be oscillatory has been proved in [1], p. 698. When (1) is of limit-circle type, a boundary condition (3) at $t = 0$ and a similar condition at $t = \infty$ determine a self-adjoint boundary value problem in $L^2(0, \infty)$ with a pure point spectrum, without a finite cluster point; [9]. The characterization in [2] of the spectra S_α for the case that (1) is of limit-point type clearly has an analogue for the spectra associated with the boundary value problems belonging to a differential equation (1) of limit-circle type. Hence, in the latter case, (7) must hold (otherwise the spectrum will have a finite cluster point λ satisfying $\lambda_1 \leqq \lambda \leqq \lambda_2$). Thus the positive part of Corollary 5 follows. The negative part follows from the proof of Corollary 1.

7. Proof of the case (3) of (*). It can be supposed that $\psi(\infty) = \infty$, otherwise (*) is trivial. For if $\psi(\infty) < \infty$, it is sufficient to choose any $q(t)$ satisfying $q(t) \to -\infty$, as $t \to \infty$. Then $q(t) + \lambda$ is negative for large t, and so a solution $x = x(t) \not\equiv 0$ of (1) has no zeros for sufficiently large t. Thus, $N(T, \lambda) = O(1)$, as $T \to \infty$, for every λ. But this is equivalent to (3), (4) and/or (5) when $\psi(\infty) < \infty$.

In order to prove the case (3) of (*), it will first be shown that there exists on $0 \leqq t < \infty$ a step-function $q^*(t)$ such that if $N(T, \lambda)$ refers to

$$y'' + (q^* + \lambda)y = 0, \tag{8}$$

rather than to (1), then (3) holds. This method for obtaining counter-examples in the theory of the differential equation (1) was introduced in [8]. By a solution of (8) is meant a function $y = y(t)$ which, on $0 \leqq t < \infty$, possesses a continuous first derivative and which, on any open interval where q^* is continuous, posssesses a continuous second derivative satisfying (8).

There is no loss of generality in supposing $\psi(0) = 0$ (for otherwise $\psi(T)$ can be suitably altered in a vicinity of $T = 0$, but the behavior of

$\psi(T)$ near $T=0$ does not enter the hypothesis or assertion of (*)). Let $0=a_0<a_1<\cdots$ denote an (unbounded) sequence of T-values satisfying

$$\psi(a_n)=n \text{ for } n=0,1,\cdots. \tag{9}$$

Define a sequence of numbers $(1<)\ \nu_0<\nu_1<\cdots$ in such a way that $\nu_n\to\infty$, as $n\to\infty$, and

$$a_n<b_n<a_{n+1}, \text{ where } b_n=a_n+\pi\nu_n^{-\frac{1}{2}} \text{ and } n=0,1,\cdots. \tag{10}$$

Introduce the abbreviation

$$\mu_n=\nu_{n+1}^5 \text{ for } n=0,1,\cdots. \tag{11}$$

It will be supposed that $\nu_0,\nu_1,\cdots$ increases so rapidly that

$$\nu_{n+1}=o(\mu_n^{\frac{1}{2}}(a_{n+1}-b_n)), \text{ as } n\to\infty. \tag{12}$$

Define $q^*=q^*(t)$ for $0\leqq t<\infty$ as follows:

$$q^*(t)=\nu_n \text{ on } a_n\leqq t<b_n; \quad q^*(t)=-\mu_n \text{ on } b_n\leqq t<a_{n+1}$$

for $n=0,1,\cdots$. Thus a solution of (8) has the form

$$y(t)=\alpha_n\cos[(\nu_n+\lambda)^{\frac{1}{2}}(t-a_n)]+\beta_n\sin[(\nu_n+\lambda)^{\frac{1}{2}}(t-a_n)], \tag{13}$$

$$y(t)=A_n\exp[(\mu_n-\lambda)^{\frac{1}{2}}(t-b_n)]+B_n\exp[-(\mu_n-\lambda)^{\frac{1}{2}}(t-b_n)], \tag{14}$$

on the respective intervals $a_n<t<b_n$, $b_n<t<a_{n+1}$, for sufficiently large n. Such a solution is determined for large t-values by fixing, say, α_K, β_K for a sufficiently large K and determining the α_n,β_n and the A_n,B_n, for $n\geqq K$, by the conditions that

$$y(t-0)=y(t+0) \text{ and } y'(t-0)=y'(t+0) \tag{15}$$

hold at the points $t=b_K, a_{K+1}, b_{K+1}, a_{K+2},\cdots$.

8. In the sequel, the following abbreviation will be used: If $c_1,c_2,\cdots$ and $d_1,d_2,\cdots$ are two sequences, the symbol $c_n=O(d_n)$, as $n\to\infty$, will signify, as usual, that there exist a constant M such that $|c_n|\leqq M|d_n|$ for all sufficiently large n. But the symbol $c_n=O_M(d_n)$ will signify $|c_n|\leqq M|d_n|$ for the specified constant M and for the specified index n (rather than for some constant M and all sufficiently large n).

Let $\lambda\neq 0$ be fixed. It will be shown that there exists a positive integer $K=K(\lambda)$, so large that $\nu_K+\lambda>0$ and that, for all $n\geqq K$,

$$\sin(\nu_n+\lambda)^{\frac{1}{2}}(b_n-a_n)=-\pi\lambda/2\nu_n+O_1(1/\nu_n^2), \tag{16}$$

(17) $\qquad (\nu_n+\lambda)^{\frac{1}{2}}(\mu_n-\lambda)^{-\frac{1}{2}} = (\nu_n/\mu_n)^{\frac{1}{2}}(1+O_{|\lambda|}(1/\nu_n)) = O_2(1/\nu_{n+1}^2),$

(18) $\qquad (\mu_n-\lambda)^{\frac{1}{2}}(\nu_{n+1}+\lambda)^{-\frac{1}{2}} = \nu_{n+1}^2(1+O_{|\lambda|}(1/\nu_{n+1})),$

and that if $0 \leqq h,\ j \leqq 2 + |\lambda|$, then for all $n \geqq K$

(19) $\qquad (1+O_h(1/\nu_n))(1+O_j(1/\nu_n))^{\pm 1} = 1 + O_{h+j+1}(1/\nu_n) = O_2(1),$

and that if C is a fixed constant, say

(20) $$C = 18/\pi\,|\lambda|,$$

then for all $n \geqq K$

(21) $$(1+O_C(1/\nu_n))(1+O_C(1/\nu_n))^{-1}\exp[-2(\mu_n-\lambda)^{\frac{1}{2}}(a_{n+1}-b_n)] = O_1(1/\nu_{n+1}).$$

The definition of b_n in (10) shows that

$$(\nu_n+\lambda)^{\frac{1}{2}}(b_n-a_n) = (\nu_n+\lambda)^{\frac{1}{2}}\pi/\nu_n^{\frac{1}{2}} = \pi(1+\lambda/\nu_n)^{\frac{1}{2}} = \pi(1+\lambda/2\nu_n+\cdots);$$

so that the left-hand side of (16) is $-\pi\lambda/2\nu_n + (\pi\lambda/2\nu_n)^3/3! + \cdots$, as $n\to\infty$. Clearly, K can be chosen so large that (16) holds for all $n \geqq K$.

The left-hand side of (17) is

$$(\nu_n/\mu_n)^{\frac{1}{2}}(1+\lambda/\nu_n)^{\frac{1}{2}}(1-\lambda/\mu_n)^{-\frac{1}{2}} = (\nu_n/\mu_n)^{\frac{1}{2}}(1+\lambda/2\nu_n+\cdots)(1+\lambda/2\mu_n+\cdots).$$

In view of the definition (11) of μ_n, the last expression is

$$(\nu_n/\mu_n)^{\frac{1}{2}}(1+O_{|\lambda|}(1/\nu_n))$$

if n is sufficiently large. Also, $\nu_n/\mu_n = \nu_n/\nu_{n+1}^5 \leqq 1/\nu_{n+1}^4$. Hence (17) holds for all $n \geqq K$ if K is sufficiently large.

Similarly, it is seen that (18) holds for all $n \geqq K$ if K is sufficiently large. It is also clear that if K is sufficiently large, then (19) is valid for all $n \geqq K$.

As to (21), the exponential factor, for large n, does not exceed $\exp[-\mu_n^{\frac{1}{2}}(a_{n+1}-b_n)]$ which, according to (12), is $o(\exp[-\nu_{n+1}])$, as $n\to\infty$. This makes obvious the existence of a K for which (21) holds when $n \geqq K$.

9. It will be shown that if β_K is chosen to be 1, and α_K is chosen to be $O_2(\beta_K/K^2)$, for example, $\alpha_K = 0$, then, in the corresponding solution (13), (14) of (8), for $n \geqq K$,

(22_n) $\qquad A_n = -\beta_n(\pi\lambda/4\nu_n)(1+O_C(1/\nu_n)),$

(23_n) $\qquad B_n = -\beta_n(\pi\lambda/4\nu_n)(1+O_C(1/\nu_n))$

and

$$(24_n) \qquad \alpha_{n+1} = A_n(1 + O_1(1/\nu_{n+1}))\exp[(\mu_{n+1} - \lambda)^{\frac{1}{2}}(a_{n+1} - b_n)],$$

$$(25_n) \qquad \beta_{n+1} = \nu_{n+1}A_n(1 + O_{|\lambda|+2}(1/\nu_{n+1}))\exp[(\mu_{n+1} - \lambda)^{\frac{1}{2}}(a_{n+1} - b_n)].$$

It can be remarked that if (24_n), (25_n) hold, then by (19),

$$(26_n) \qquad \alpha_{n+1} = O_2(\beta_{n+1}/\nu_{n+1}^2).$$

The proof of (22_n)-(25_n) for $n \geqq K$ will be by mathematical induction on n. Since the passage from n to $n+1$ will depend on (15) and (26_n), it is clear from the choice of α_K, β_K that a direct verification of (22_K)-(25_K) need not be made.

Let $n(\geqq K)$ be fixed and suppose that (22_n)-(25_n) hold; in particular, that (26_n) holds. The relations (15), where $t = b_{n+1}$, give, by virtue of (13), (14),

$$A_{n+1} + B_{n+1} = \alpha_{n+1}\cos + \beta_{n+1}\sin,$$

$$A_{n+1} - B_{n+1} = (\nu_{n+1} + \lambda)^{\frac{1}{2}}(\mu_{n+1} - \lambda)^{-\frac{1}{2}}\{-\alpha_{n+1}\sin + \beta_{n+1}\cos\},$$

where the argument of sin, cos is $(\nu_{n+1} + \lambda)^{\frac{1}{2}}(b_{n+1} - a_{n+1})$. By (16), (17) and (26_n), these equations take the form

$$A_{n+1} + B_{n+1} = O_2(\beta_{n+1}/\nu_{n+1}^2) + \beta_{n+1}(-\pi\lambda/2\nu_{n+1} + O_1(1/\nu_{n+1}^2)),$$

$$A_{n+1} - B_{n+1} = O_2(1/\nu_{n+2}^2)\{O_2(\beta_{n+1}/\nu_{n+1}^2) + O_1(\beta_{n+1})\}.$$

Since $1/\nu_{n+2}^2 \leqq 1/\nu_{n+1}^2 < 1$, it follows that

$$A_{n+1} = -\beta_{n+1}(\pi\lambda/4\nu_{n+1}) + \tfrac{1}{2}O_{2+1+4+2}(\beta_{n+1}/\nu_{n+1}^2).$$

Hence, (22_{n+1}) follows from the definition (20) of C. If the last two simultaneous equations are solved for B_{n+1}, it is seen that (23_{n+1}) holds.

The relations (15), where $t = a_{n+2}$, give

$$\alpha_{n+2} = A_{n+1}\exp[\cdot\ \cdot\ \cdot] + B_{n+1}\exp(-[\cdot\ \cdot\ \cdot]),$$

$$\beta_{n+2} = (\mu_{n+1} - \lambda)^{\frac{1}{2}}(\nu_{n+2} + \lambda)^{-\frac{1}{2}}\{A_{n+1}\exp[\cdot\ \cdot\ \cdot] - B_{n+1}\exp(-[\cdot\ \cdot\ \cdot])\},$$

where

$$(27) \qquad [\cdot\ \cdot\ \cdot] = (\mu_{n+1} - \lambda)^{\frac{1}{2}}(a_{n+2} - b_{n+1}).$$

In view of (18), these equations can be written as

$$(28) \quad \alpha_{n+2} = A_{n+1}\{1 + (B_{n+1}/A_{n+1})\exp(-2[\cdot\ \cdot\ \cdot])\}\exp[\cdot\ \cdot\ \cdot],$$

$$(29)\quad \beta_{n+2} = \nu_{n+2}{}^2(1 + O_{|\lambda|}(1/\nu_{n+2}))A_{n+1} \times \{1 - (B_{n+1}/A_{n+1})\exp(-2[\cdot\cdot\cdot])\}\exp[\cdot\cdot\cdot].$$

Since (22_{n+1}), (23_{n+1}) have just been verified, it follows that

$$B_{n+1}/A_{n+1} = (1 + O_C(1/\nu_{n+1}))(1 + O_C(1/\nu_{n+1}))^{-1}.$$

Thus (24_{n+1}) is a consequence of (28) and (21), while (25_{n+1}) follows from (29), (21), and (18). This completes the induction and so (22_n)-(25_n) hold for all $n \geqq K$.

10. Without affecting the preceding considerations, it can be supposed that K has been chosen so large that

$$(30)\quad 1 + O_h(1/\nu_n) > \tfrac{1}{2} > 0 \text{ for } n \geqq K \text{ if } h = 1,\ C \text{ or } |\lambda| + 2.$$

It follows from (22_n), (23_n) that

$$(31)\qquad A_n \sim B_n, \text{ as } n \to \infty,$$

and that A_n, B_n have the same sign at each $n \geqq K$. Also, (22_{n+1}) and (25_n) show that

$$(32)\quad A_{n+1} = -(\pi\lambda/4\nu_{n+1})A_n(1 + O_C(1/\nu_{n+1}))(1 + O_{|\lambda|+2}(1/\nu_{n+1}))\exp[\cdot\cdot\cdot],$$

where $[\cdot\cdot\cdot]$ is given by (27).

Since A_n and B_n, at each $n \geqq K$, have the same sign, it follows from (14) that $y(t)$ has no zero on $b_n \leqq t \leqq a_{n+1}$. If $\lambda > 0$, then A_n and A_{n+1} are of opposite signs, by (30), (32). It follows, therefore, that $y(t)$ has an odd number of zeros on $a_{n+1} < t < b_{n+1}$. But (13) and (10) show that $y(t)$ has on this last interval at most $1 + (\nu_{n+1} + \lambda)^{\frac{1}{2}}/\nu_{n+1}{}^{\frac{1}{2}}$ zeros. Since $1 + (\nu_{n+1} + \lambda)^{\frac{1}{2}}/\nu_{n+1}{}^{\frac{1}{2}} \to 2$, as $n \to \infty$, it follows that $y(t)$ has, for large n, exactly one zero on $a_{n+1} < t < b_{n+1}$. Thus the second part of (3) is a consequence of the definition, (9), of a_n.

If $\lambda < 0$, then A_n and A_{n+1} are of the same sign, and so $y(t)$ has an even number of zeros on $a_{n+1} < t < b_{n+1}$. But, on this interval, $y(t)$ has at most $1 + (\nu_{n+1} + \lambda)^{\frac{1}{2}}/\nu_{n+1}{}^{\frac{1}{2}}$ zeros. Since $1 + (\nu_{n+1} + \lambda)^{\frac{1}{2}}/\nu_{n+1}{}^{\frac{1}{2}} < 2$ when $\lambda < 0$, it follows that $y(t)$ has no zero on $a_{n+1} < t < b_{n+1}$. Thus the first part of (3) is proved.

11. This proves the case (3) of (*) if $q(t)$ is allowed to be a step-function, $q^*(t)$. It will be shown that $q^*(t)$ can be modified so as to become a continuous function $q(t)$, for which (3) remains valid.

Since every solution $y = y(t)$ of (8) has an absolutely continuous first derivative, one has $d(y^2 + y'^2) = 2y'(y + y'')dt = -2yy'(q^* + \lambda - 1)dt$. Thus $| d(y^2 + y'^2) | \leqq (y^2 + y'^2)(| q^* | + | \lambda | + 1)dt$; and so $y^2 + y'^2 = O(\exp \int_0^t (| q^*(s) | + | \lambda | + 1)ds)$, as $t \to \infty$. In particular,

$$(33) \qquad | y(t) | = O(\Omega(t)), \text{ as } t \to \infty,$$

where $\Omega(t) = \exp \frac{1}{2} \int_0^t (| q^*(s) | + s)ds$. The point in the inequality (33) is that $y(t)$, being a solution of (8), depends on λ but the majorant $\Omega(t)$ is independent of λ.

Starting with $q^*(t)$ construct a continuous function $q(t)$ on $0 \leqq t < \infty$ by letting $q(t) = q^*(t)$ except on small intervals, say $(b_0, b_0 + \epsilon_0)$, $(a_1 - \epsilon_1, a_1)$, $(b_1, b_1 + \epsilon_1)$, $(a_2 - \epsilon_2, a_2), \cdots$, where $q(t)$ is linear. Let

$$(34) \qquad \epsilon(t) = q(t) - q^*(t).$$

It is clear that if the numbers $\epsilon_0, \epsilon_1, \cdots$ tend to 0 sufficiently rapidly, then

$$(35) \qquad \int^{\infty} | \epsilon(t) | \Omega^2(t) dt < \infty.$$

Let $y = y_1(t) = y_1(t, \lambda)$ and $y = y_2(t) = y_2(t, \lambda)$ be a pair of solutions of (8) satisfying the Wronskian condition $y'_1 y_2 - y_1 y'_2 \equiv 1$. If (1) is written as $x'' + (q^* + \lambda)x = (q^* - q)x$, it is seen that every solution $x = x(t) = x(t, \lambda)$ of (1), and its derivative, can be written in the form

$$(36) \qquad x = p_2 y_1 - p_1 y_2 \text{ and } x' = p_2 y'_1 - p_1 y'_2$$

where, for $j = 1, 2$,

$$p_j = p_j(t, \lambda) = \gamma_j + \int_0^t (q^*(s) - q(s)) y_j(s) x(s) ds,$$

and γ_1. γ_2 are constants. The results of [10], pp. 261-268, show that (33) and (35) imply that $c_j = \lim p_j(t, \lambda)$, as $t \to \infty$, exist for $j = 1, 2$; furthermore, $c_1 = c_2 = 0$ only if $x(t) \equiv 0$. (In [10], q and q^* are assumed to be continuous, but it is clear that the discontinuities of q^* do not affect the arguments there.)

Let $x(t) = x(t, \lambda) \not\equiv 0$; so that at least one the numbers c_1, c_2 is not zero. For the sake of concreteness, let $c_1 \neq 0$. Then the Wronskian condition on y_1, y_2 and (36) imply $xy'_1 - x'y_1 = -p_1 \rightarrow -c_1$, as $t \rightarrow \infty$. Consequently, for large t-values, x and y_1 do not vanish simultaneously and $(\arctan y_1/x)'$ does not change sign; and so, the zeros of y_1 and x separate each other; cf. also [7]. Thus (3) holds for (1), since it holds for (8). This completes the proof of the case (3) of (*).

12. Proof of the case (4) of (*). This part of (*) is proved in a manner similar to that of case (3). The main differences will be indicated. In (10), let the numbers b_n be defined by

$$b_n = a_n + 3\pi/2\nu_n^{\frac{1}{2}} \qquad \text{for } n = 0, 1, \cdots. \tag{37}$$

Then (16) can be replaced by

$$\sin(\nu_n + \lambda)^{\frac{1}{2}}(b_n - a_n) = -1 + O_M(1/\nu_n^2), \tag{38}$$

where, say, $M = M(\lambda) = 5\lambda^2 (> (3\pi\lambda/4)^2/2!)$. By a suitable choice of the integer K and the constant D (depending on λ), it can be shown that the corresponding equation (8) has a solution (13), (14), where, for $n \geqq K$,

$$A_n = -(\beta_n/2)(1 + O_D(1/\nu_n)), \quad B_n = -(\beta_n/2)(1 + O_D(1/\nu_n)),$$

while (24_n), (25_n) hold.

In this case, A_n and A_{n+1} are of opposite sign for all λ. Thus, for the differential equation (8), the relation (4) can be verified as in § 10. The passage from (8) to (1) is the same as in § 11.

13. Proof of the case (5) of (*). This can be proved by another choice of b_n, namely,

$$b_n = a_n + 2\pi/\nu_n^{\frac{1}{2}} \qquad \text{for } n = 0, 1, \cdots. \tag{39}$$

Then (16) becomes

$$\sin(\nu_n + \lambda)^{\frac{1}{2}}(b_n - a_n) = \pi\lambda/\nu_n + O_1(1/\nu_n^2). \tag{40}$$

By the procedure of § 9, it can be shown that the corresponding equation (8) has a solution (13), (14), where, for $n \geqq K$,

$$A_n = \beta_n(\pi\lambda/2\nu_n)(1 + O_C(1/\nu_n)), \quad B_n = \beta_n(\pi\lambda/2\nu_n)(1 + O_C(1/\nu_n)),$$

while (24_n), (25_n) hold. Clearly, the proof can be finished by the procedures of §§ 10-11.

This completes the proof of (*).

14. Proof of (3 bis) and (5 bis). Return to the proof of the case (3) in § 9. It will be shown that if $\lambda = 0$ and K is a suitably chosen integer, then (8) has a solution (13), (14) where, for $n \geqq K$,

$$(41_n)\qquad \begin{aligned} A_n &= -\tfrac{1}{2}\beta_n(\nu_n^{-2} + \nu_{n+1}^{-2})(1 + O_8(1/\nu_n)),\\ B_n &= -\tfrac{1}{2}\beta_n(\nu_n^{-2} - \nu_{n+1}^{-2})(1 + O_8(1/\nu_n)),\end{aligned}$$

and

$$(42_n)\qquad \begin{aligned} \alpha_{n+1} &= A_n(1 + O_1(1/\nu_{n+1}))\exp[\mu_n^{\frac{1}{2}}(a_{n+1} - b_n)],\\ \beta_{n+1} &= \nu_{n+1}^2 A_n(1 + O_1(1/\nu_{n+1}))\exp[\mu_n^{\frac{1}{2}}(a_{n+1} - b_n)].\end{aligned}$$

The proof of these formulae, by induction, will be similar to the proof of (22_n)-(25_n). The relations (42_n) imply, if n is sufficiently large,

$$(43_n)\qquad \alpha_{n+1} = (\beta_{n+1}/\nu_{n+1}^2)(1 + O_4(1/\nu_{n+1})).$$

Let K be a positive integer. to be fixed below. Let α_K, β_K be chosen so that (43_{K-1}) holds.

Assume (41_n) and (42_n). The equations for A_{n+1}, B_{n+1}, in the formula lines following (26) become for $\lambda = 0$

$$A_{n+1} + B_{n+1} = -\alpha_{n+1}, \qquad A_{n+1} - B_{n+1} = -\nu_{n+2}^{-2}\beta_{n+1}.$$

Thus, by (43_n),

$$\begin{aligned} A_{n+1} &= -\tfrac{1}{2}\beta_{n+1}(\nu_{n+2}^{-2} + \nu_{n+2}^{-2}) + O_2(\beta_{n+1}/\nu_{n+1}^3),\\ B_{n+1} &= -\tfrac{1}{2}\beta_{n+1}(\nu_{n+2}^{-2} - \nu_{n+2}^{-2}) + O_2(\beta_{n+1}/\nu_{n+1}^3),\end{aligned}$$

and so (41_{n+1}) holds.

The equations for α_{n+2}, β_{n+2} become identical with those preceding (27), where $\lambda = 0$. These can be written in the form

$$\begin{aligned} \alpha_{n+2} &= A_{n+1}\{1 + (B_{n+1}/A_{n+1})\exp(-2[\cdot\ \cdot\ \cdot])\}\exp[\cdot\ \cdot\ \cdot],\\ \beta_{n+2} &= \nu_{n+2}^2 A_{n+1}\{1 - (B_{n+1}/A_{n+1})\exp(-2[\cdot\ \cdot\ \cdot])\}\exp[\cdot\ \cdot\ \cdot],\end{aligned}$$

where $[\cdot\ \cdot\ \cdot]$ is given by (27) when $\lambda = 0$. By (41_{n+1}), which has just been

verified, $B_{n+1}/A_{n+1} = O_1(1)(1 + O_8(1/\nu_{n+1}))(1 + O_8(1/\nu_{n+1}))^{-1}$. Clearly, if K is fixed sufficientiy large, then (42_{n+1}) holds.

The proof can now be completed as in § 10. The case $\lambda = 0$ of (5) is similar.

15. Proof of the Remark on (3_0) and (5_0). Let a_n, b_n, ν_n, μ_n have the same significance as in (9), (10), (11), (12). Define $\delta_0, \delta_1, \cdots$ by placing $\delta_n = \pm 1/\nu_n^{\frac{1}{2}}$, where the $\pm$ is a function of n to be given below. Let $q^*(t)$ be defined, for $0 \leqq t < \infty$, as follows:

$$q^*(t) = \nu_n + \delta_n \text{ on } a_n \leqq t < b_n; \qquad q^*(t) = -\mu_n + \delta_n \text{ on } b_n \leqq t < a_{n+1}$$

for $n = 0, 1, \cdots$. Let (13′), (14′) denote the equation which result if λ is replaced by $\lambda + \delta_n$ in (13), (14). Then the solutions of (8), for large t, are given by (13′), (14′). It is clear from §§ 8-9 that if $\lambda \neq 0$, then (8) has a solution (13′), (14′) where, for large n, the coefficients satisfy the equations which result if $\lambda + \delta_n$, $\lambda + \delta_{n+1}$ are written in place of λ in (22_n)-(23_n), (24_n)-(25_n), respectively. Thus, by the arguments of § 10, the asymptotic formulae (3) hold if $N(T, \lambda)$ refers to (8).

It remains to consider (3_0). It will be shown that (8), where $\lambda = 0$, has a solution (13′). (14′) with coefficients satisfying

$$(44_n) \qquad \begin{aligned} A_n &= -\beta_n(\pi\delta_n/4\nu_n)(1 + O_M(1/\nu_n^{\frac{1}{2}})), \\ B_n &= -\beta_n(\pi\delta_n/4\nu_n)(1 + O_M(1/\nu_n^{\frac{1}{2}})) \end{aligned}$$

and

$$(45_n) \qquad \begin{aligned} \alpha_{n+1} &= A_n(1 + O_3(1/\nu_{n+1}))\exp[(\mu_{n+1} - \delta_{n+1})^{\frac{1}{2}}(a_{n+1} - b_n)], \\ \beta_{n+1} &= \nu_{n+1}{}^2 A_n(1 + O_3(1/\nu_{n+1}))\exp[(\mu_{n+1} - \delta_{n+1})^{\frac{1}{2}}(a_{n+1} - b_n)], \end{aligned}$$

for large n, where $M = 18/\pi$.

The equations (45_n) imply (26_n), if n is sufficiently large. Let K be a positive integer, to be fixed below. Let α_K, β_K be chosen so that (26_{K-1}) holds and assume that, for a fixed $n (\geqq K)$, the relations (44_n), (45_n) hold.

The equations for A_{n+1}, B_{n+1} become those in the formula lines following (26_n), where λ must be replaced by δ_{n+1}. Since (16) and (17) hold for large n, with $\lambda = \delta_n$, the equations A_{n+1}, B_{n+1} become identical with those in the second pair of formula lines following (26_n), with the modification that λ is δ_{n+1}. Hence, the definition of δ_n shows that (44_{n+1}) holds (if K is sufficiently large).

Since (18) can be improved, for $\lambda = \delta_n$ and large n, to

$$(\mu_n - \delta_n)^{\frac{1}{2}}(\nu_{n+1} + \delta_{n+1})^{-\frac{1}{2}} = \nu_{n+1}{}^2(1 + O_1(1/\nu_{n+1})),$$

the deduction of (24_n), (25_n) above shows that (45_{n+1}) is a consequence of (44_{n+1}) if n is sufficiently large.

Let K be so large that (8), where $\lambda = 0$, has a solution $y = y(t)$ for which (44_n), (45_n) hold for $n \geqq K$. The arguments of § 10 show that, for large n, $y(t)$ has one or no zero for $a_n \leqq t < a_{n+1}$ according as the $+$ or $-$ sign holds in the definition of δ_n. It is clear that, for a given $\phi(T)$, the $\pm$ signs (as function of n) can be chosen so that (3_0) holds if $N(T, 0)$ refers to (8). Since the passage from (8) to (1) is the same as in § 11, the case (3_0) is proved.

Clearly, the case (5_0) is proved similarly with the choice $b_n = a_n + 2\pi/\nu_n{}^{\frac{1}{2}}$.

16. Proof of (**). It can be supposed that the given set S is not empty. Otherwise it is sufficient to choose $q(t)$ so that $q(t) \to -\infty$, as $t \to \infty$. Then $N(T, \mu) - N(T, \lambda) = O(1)$, as $T \to \infty$, for $-\infty < \lambda < \mu < \infty$.

Clearly, § 11 implies that it is sufficient to prove the existence of a function $q^*(t)$, $0 \leqq t < \infty$, which is a step-function with discontinuities clustering only at $t = \infty$ and for which the assertion (**) holds if (1) is replaced by (8).

Let $\tau^1, \tau^2, \cdots$ be a sequence of points of the λ-set S (allowing repetitions) with the property that every point of S is either a point or a cluster point of the sequence. Let $\tau_1, \tau_2, \cdots$ denote the sequence $\tau^1; \tau^1, \tau^1, \tau^2, \tau^2; \tau^1, \tau^1, \tau^1, \tau^2, \tau^2, \tau^2, \tau^3, \tau^3, \tau^3; \tau^1, \cdots$; so that, for a given k, there exist arbitrarily large N and M such that $\tau_n = \tau^k$ for $n = N, N+1, \cdots, M$.

Let $\psi(t)$ be a continuous, monotone function satisfying $\psi(0) = 0$ and $\psi(\infty) = \infty$, and let (9), (10), (11) and (12) hold. In addition, it can be supposed that $\nu_0, \nu_1, \cdots$ increases so rapidly that, as $n \to \infty$,

$$\sigma_n{}^3 = o(\nu_n), \text{ where } \sigma_n = n + \max(|\tau_1|, \cdots, |\tau_n|).$$

Define $q^* = q^*(t)$, for $0 \leqq t < \infty$, as follows:

$$q^*(t) = \nu_n - \tau_n \text{ for } a_n \leqq t < b_n \text{ and } q^*(t) = -\mu_n - \tau_n \text{ for } b_n \leqq t < a_{n+1},$$

where $n = 0, 1, \cdots$. Corresponding to (13), (14), a solution of (8) has, for large n, the form

$$(46) \quad y(t) = \alpha_n \cos[(\nu_n + \lambda - \tau_n)^{\frac{1}{2}}(t - a_n)] + \beta_n \sin[(\nu_n + \lambda - \tau_n)^{\frac{1}{2}}(t - a_n)]$$

$$(47)\quad y(t) = A_n \exp[(\mu_n - \lambda + \tau_n)^{\frac{1}{2}}(t - b_n)] \\ + B_n \exp[-(\mu_n - \lambda + \tau_n)^{\frac{1}{2}}(t - b_n)]$$

for $a_n \leqq t < b_n$, $b_n \leqq t < a_{n+1}$, respectively.

Let $\epsilon > 0$ and λ be fixed. A slight modification of the procedures in § 8 shows that there exists a positive integer $K = K(\lambda, \epsilon)$ so large that for $n \geqq K$ one has $\nu_n + \lambda - \tau_n > 0$ and

$$(48)\quad \sin(\nu_n + \lambda - \tau_n)^{\frac{1}{2}}(b_n - a_n) = -\pi(\lambda - \tau_n)/2\nu_n + O_1(1/\nu_n^2),$$

$$(49)\quad (\nu_n + \lambda - \tau_n)^{\frac{1}{2}}(\mu_n - \lambda + \tau_n)^{-\frac{1}{2}} = (\nu_n/\mu_n)^{\frac{1}{2}}(1 + O_1(\sigma_n/\nu_n)) = O_2(1/\nu_{n+1}^2),$$

$$(50)\quad (\mu_n - \lambda + \tau_n)^{\frac{1}{2}}(\nu_{n+1} + \lambda - \tau_n)^{-\frac{1}{2}} = \nu_{n+1}^2(1 + O_1(\sigma_{n+1}/\nu_{n+1})),$$

$$(51)\quad (1 + O_1(1/\nu_n))(1 + O_1(\sigma_n/\nu_n)) = 1 + O_1((2 + \sigma_n)/\nu_n) = O_2(1),$$

$$(52)\quad (1 + O_h(1/\nu_n)(1 + O_j(1/\nu_n)) = (1 + O_{h+j+1}(1/\nu_n)), 0 \leqq h, j \leqq 9,$$

and, in addition, for

$$(53)\quad C = 18/\pi\epsilon,$$

one has

$$(54)\quad (1 + O_C(1/\nu_n))(1 + O_C(1/\nu_n))^{-1}\exp(-2\{\cdot\cdot\cdot\}) = O_1(1/\nu_{n+1}),$$

where

$$(55)\quad \{\cdot\cdot\cdot\} = (\mu_n - \lambda + \tau_n)^{\frac{1}{2}}(a_{n+1} - b_n).$$

These relations and the arguments of § 9 show that if $N \geqq K(\lambda, \epsilon)$ and

$$(56)\quad |\lambda - \tau_n| \geqq \epsilon > 0 \text{ for } n = N, N+1, \cdots, M,$$

then (8) has a solution $y = y(t)$ for $a_N \leqq t \leqq b_{M+1}$, given by (46), (47), where $\alpha_N = O_2(\beta_N/\nu_N^2)$ and, for $n = N, N+1, \cdots, M$,

$$(57_n)\quad A_n = -(\beta_n\pi(\lambda - \tau_n)/4\nu_n)(1 + O_C(1/\nu_n)),$$

$$(58_n)\quad B_n = -(\beta_n\pi(\lambda - \tau_n)/4\nu_n)(1 + O_C(1/\nu_n))$$

and

$$(59_n)\quad \alpha_{n+1} = A_n(1 + O_1(1/\nu_{n+1}))\exp\{\cdot\cdot\cdot\},$$

$$(60_n)\quad \beta_{n+1} = \nu_{n+1}^2 A_n(1 + O_1((2 + \sigma_{n+1})/\nu_{n+1}))\exp\{\cdot\cdot\cdot\}.$$

If K is so large that $|O_1((2 + \sigma_n)/\nu_n)| < \frac{1}{2}$, $|O_C(1/\nu_n)| < \frac{1}{2}$ and $1 + (\nu_{n+1} + \lambda - \tau_{n+1})^{\frac{1}{2}}/\nu_{n+1}^{\frac{1}{2}} \leqq 5/2 < 3$ whenever $n \geqq K$, then the methods of

§ 10 show that (57_n)-(60_n) imply that $y = y(t)$ has exactly one or no zero on the intervals $b_n \leqq t \leqq b_{n+1}$ according as $\lambda > \tau_n$ or $\lambda < \tau_n$, where $n = N, N+1, \cdots, M$.

Let $\lambda < \mu$ and suppose that there exists an $\epsilon > 0$ such that, for all sufficiently large n,

$$(61) \qquad | \lambda - \tau_n | \geqq \epsilon > 0 \text{ and } | \mu - \tau_n | \geqq \epsilon > 0.$$

Then $N(T, \mu) - N(T, \lambda)$ is unbounded or bounded according as there is or is not at least one value of n (hence infinitely many) for which $\lambda < \tau_n < \mu$. This proves the assertion (**), for (8), in the case in which S has no point on the closed interval $[\lambda, \mu]$.

If the open interval (λ, μ) contains points of S, the structure of the sequence $\tau_1, \tau_2, \cdots$ shows that there exist an $\epsilon > 0$ and a sequence of increasing integers $N_1 < M_1 < N_2 < M_2 < \cdots$ such that $M_n - N_n \to \infty$, as $n \to \infty$, and $\lambda + \epsilon < \tau_n < \mu - \epsilon$ if $n = N_k, N_k + 1, \cdots, M_k$, where $k = 1, 2, \cdots$. Thus, (56) holds for $N = N_k$ and $M = M_k$. Hence (8) has on the interval $b_N \leqq t \leqq b_M$, where $N = N_k$ and $M = M_k$, a solution with no zeros; while if λ is replaced by μ, the resulting differential equation (8) has a solution with exactly $M - N$ zeros. Thus the contribution of $b_N \leqq t \leqq b_M$ to $N(T, \mu) - N(T, \lambda)$, for $T \geqq b_M$, is at least $(M - N) - 2$. Consequently, $M_n - N_n \to \infty$, as $n \to \infty$, implies that $N(T, \mu) - N(T, \lambda) \to \infty$, as $T \to \infty$. This completes the proof of (**).

17. Proof of the Remark following (**). If S is bounded from below, the above proof of (**) gives an example in which (1) is non-oscillatory for λ less than every number in S. In order to obtain an example in which (1) is oscillatory for every λ, it is sufficient to repeat the above construction, with b_n in (10) replaced by $b_n = a_n + 2\pi/\nu_n^{\frac{1}{2}}$.

APPENDIX.*

In the paper "Oscillatory and non-oscillatory differential equations" by Wintner and the author, this Journal, vol. 71 (1949), p. 646, the question is raised as to whether or not the condition

$$(62) \qquad q(t) \text{ tends monotonously to } \infty, \text{ as } t \to \infty,$$

implies that the essential spectrum of (1) is the entire λ-axis, when (2) is of limit-point type. It was proved *loc. cit.* that the answer is in the affirmative

* Recieved May 5, 1951.

when the growth of q is sufficiently smooth. The object of this Appendix is to adapt the procedures employed above to prove that, in general, the answer is in the negative.

There exist on $0 \leqq t < \infty$ continuous functions $q(t)$ having the properties that (62) holds, that (1) is of limit-point type, and that the spectrum S_α is a pure point spectrum clustering at, and only at, ∞ and $-\infty$.

This assertion can be considered as a refinement of Corollaries 1 and 2 in § 2. The considerations of § 11 show that it is sufficient to construct a step-function $q^*(t)$ having all of the desired properties, except that of continuity. It turns out that the example in the Appendix of [6], pp. 211-212, is of this type. In this particular example, the function q (or q^*) satisfies

$$\int^{\infty} dt/q^{\frac{1}{2}} < \infty.$$

It will remain an open question whether or not this holds for every example.

The function $q^*(t)$ in the Appendix of [6] is defined on $0 \leqq t < \infty$ by

$$q^*(t) = \nu_n \text{ if } a_n \leqq t < a_{n+1} \tag{63}$$

and

$$a_1 = 0 \text{ and } a_{n+1} - a_n = 2\pi/\nu_n^{\frac{1}{2}} \text{ for } n = 1, 2, \cdots, \tag{64}$$

finally

$$\nu_n = n^2; \tag{65}$$

so that $a_n \to \infty$, while $q^*(t)$ satisfies (62). This function makes (8) of limit-point type, by [6], pp. 211-212. It will be shown that, for every λ, (8) has a solution which vanishes exactly twice on $a_n \leqq t < a_{n+1}$ for large n; in particular, $N(T, \lambda) - N(T, \mu) = O(1)$, as $T \to \infty$, for $-\infty < \lambda < \mu < \infty$. This implies, by [2], the statement concerning S_α.

Every solution of (8) is for large t of the form

$$y(t) = \alpha_n \cos (\nu_n + \lambda)^{\frac{1}{2}}(t - a_n) + \beta_n \sin (\nu_n + \lambda)^{\frac{1}{2}}(t - a_n), \tag{66}$$

where some pair of constants α_K, β_K can be chosen arbitrarily and $\alpha_{K+1}, \beta_{K+1}, \alpha_{K+2}, \cdots$ are chosen so as to satisfy (15). The relations between α_n, β_n and $\alpha_{n+1}, \beta_{n+1}$ are given by

$$\begin{aligned} \alpha_{n+1} &= \alpha_n \cos[\cdot\cdot\cdot] + \beta_n \sin[\cdot\cdot\cdot], \\ \beta_{n+1} &= (\nu_n + \lambda)^{\frac{1}{2}}(\nu_{n+1} + \lambda)^{-\frac{1}{2}}\{-\alpha_n \sin[\cdot\cdot\cdot] + \beta_n \cos[\cdot\cdot\cdot]\}, \end{aligned} \tag{67}$$

where the argument $[\cdot\cdot\cdot]$ of sin and cos is, by (64), $2\pi(\nu_n+\lambda)^{\frac{1}{2}}\nu_n^{-\frac{1}{2}}$, which is of the form $2\pi(1+\lambda/2\nu_n+\cdots)$, if λ fixed and $n\rightarrow\infty$. Consequently, if n is sufficiently large, $\cos[\cdot\cdot\cdot]=1+O_M(\lambda^2/\nu_n^2)$ and $\sin[\cdot\cdot\cdot]=O_M(\lambda/\nu_n)$, where $M=\pi^2$, and $(\nu_n+\lambda)^{\frac{1}{2}}(\nu_{n+1}+\lambda)^{-\frac{1}{2}}\leqq 1-1/2\nu_n^{\frac{1}{2}}$, by (65). Hence, (67) becomes

$$(68)\qquad \begin{aligned} &\alpha_{n+1}=\alpha_n(1+O_M(\lambda^2/\nu_n^2))+O_M(\lambda\beta_n/\nu_n),\\ &|\beta_{n+1}|\leqq(1-1/2\nu_n^{\frac{1}{2}})(O_M(\lambda\alpha_n/\nu_n)+|\beta_n|). \end{aligned}$$

If $\alpha_n\neq 0$, put $c_n=\beta_n/\alpha_n$. It will be shown, by induction, that if the integer $K=K(\lambda)$ is sufficiently large, $\alpha_K\neq 0$ and $c_K=O_M(\lambda/\nu_{K-1})$, then $\alpha_{n+1}\neq 0$, and the relation

$$(69_{\rm n})\qquad c_{n+1}=O_M(\lambda\sum_{k=K-1}^{n}1/\nu_k)\leqq C,\ \text{where } C=C(K,\lambda)$$

is a constant (cf. (65)), holds for $n=K-1,K,\cdots$. Assume $\alpha_n\neq 0$ and that $(69_{\rm n-1})$ holds, where $n(\geqq K)$ is fixed. Then the first equation of (68) gives

$$(70_{\rm n})\qquad \alpha_{n+1}=\alpha_n(1+O_{2MC}(\lambda/\nu_n)),\ \text{where } 1+O_{2MC}(\lambda/\nu_n)>0,$$

if K is sufficiently large. In particular, $\alpha_{n+1}\neq 0$. The second equation of (68) shows that

$$|c_{n+1}|\leqq(1-1/2\nu_n^{\frac{1}{2}})(1+O_{3MC}(\lambda/\nu_n))(M|\lambda|/\nu_n+|c_n|).$$

Since the product of the first two factors on the right does not exceed 1 if K is sufficiently large, $(69_{\rm n})$ follows from $(69_{\rm n-1})$. This completes the induction.

Consequently, $(70_{\rm n})$ holds for $n=K,K+1,\cdots$. Since $y(a_n)=\alpha_n$ by (66), it follows that $y(a_n)$ and $y(a_{n+1})$ are of the same sign, so that $y(t)$ vanishes an even number of times on $a_n<t<a_{n+1}$. That $y(t)$, for large t, vanishes exactly twice on $a_n\leqq t<a_{n+1}$, follows by the arguments of § 10. This completes the proof of the italicized assertion.

PARIS, FRANCE.

REFERENCES.

[1] P. Hartman, "Differential equations with non-oscillatory eigenfunctions," *Duke Mathematical Journal*, vol. 15 (1948), pp. 697-709.

[2] ———, "A characterization of the spectra of one-dimensional wave equations," *American Journal of Mathematics*, vol. 71 (1949), pp. 915-920.

[3] ——— and C. R. Putnam, "The least cluster point of the spectrum of boundary value problems," *ibid.*, vol. 70 (1948), pp. 849-855.

[4] ——— and A. Wintner, "An oscillation theorem for continuous spectra," *Proceedings of the National Academy of Sciences*, vol. 33 (1947), pp. 376-379.

[5] ——— and A. Wintner, "On the orientation of unilateral spectra," *American Journal of Mathematics*, vol. 70 (1948), pp. 309-316.

[6] ——— and A. Wintner, "A criterion for the non-degeneracy of the wave equation," *ibid.*, vol. 71 (1949), pp. 206-213.

[7] W. Leighton, "On self-adjoint differential equations of second order," *Proceedings of the National Academy of Sciences*, vol. 35 (1949), pp. 656-657.

[8] H. Milloux, "Sur l'équation differentielle $x'' + A(t)x = 0$," *Prace Matematyczno-Fizyczne*, vol. 41 (1934), pp. 39-54.

[9] H. Weyl, "Ueber gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen," *Mathematische Annalen*, vol. 68 (1910), pp. 222-269.

[10] A. Wintner, "Asymptotic integrations of the adiabatic oscillator," *American Journal of Mathematics*, vol. 69 (1947), pp. 251-272.

[11] ———, "On Dirac's theory of continuous spectra," *Physical Review*, vol. 73 (1948), pp. 781-785.

BOOLEAN ALGEBRAS WITH OPERATORS.*

By BJARNI JÓNSSON and ALFRED TARSKI.

PART II **

Section 4.

Representation Theorems for Relation Algebras.

Relation algebras are abstract algebraic systems characterized by means of a number of simple postulates which prove to be satisfied if we take the elements of the algebra to be binary relations, and the fundamental operations of this algebra to be set-theoretical operations of addition and multiplication together with relative multiplication and conversion. The relation algebra actually formed by binary relations and the operations just mentioned will be referred to as proper relation algebras. The natural representation problem for relation algebras is the problem whether every relation algebra is isomorphic to a proper relation algebra. It has recently been shown that in general the solution of this problem is negative.[14] On the other hand, we shall see in Theorem 4.22 that every relation algebra has at least a "weak" natural representation in which all the operations except the Boolean multiplication have their natural meaning. In some further theorems, e. g., 4.29 and 4.32, we shall obtain a positive solution of the natural representation problem for certain classes of relation algebras which, however, are of a rather special nature.

When studying abstract relation alegbras it is useful to bear in mind that various notions of the general theory of these algebras take on a familiar meaning and various results can easily be anticipated when applied to proper relation algebras.

DEFINITION 4.1. *An algebra*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', \smile \rangle$$

* Received March 13, 1950.

** Part I appeared in vol. 73 (1951), pp. 891-939, of this Journal.

[14] Cf. Lyndon [1]. The numbers in brackets refer to the bibliography at the beginning of Part I of this paper.

(*where* $+$, $\cdot$, *and* ; *are operations on* A^2 *to* A, $^\smile$ *is an operation on* A *to* A, *and* 0, 1, *and* 1' *are elements of* A) *is called a relation algebra if the following conditions are satisfied:*

(i) $\langle A, +, 0, \cdot, 1\rangle$ *is a Boolean algebra.*

(ii) $(x\,;y)\,;z = x\,;(y\,;z)$ *for any* $x, y, z \in A$.

(iii) $1'\,;x = x = x\,;1'$ *for every* $x \in A$.

(iv) *The formulas* $(x\,;y)\cdot z = 0$, $(x^\smile;z)\cdot y = 0$, *and* $(z;y^\smile)\cdot x = 0$ *are equivalent for any* $x, y, z \in A$.

The operation ; is referred to as relative multiplication, the operation $^\smile$ *as conversion, and the element* 1' *as the identity element.*[15]

In view of condition (ii) of this definition we shall in general, when speaking of relation algebras, omit parentheses in expressions like

$$(x;y);z \text{ and } x;(y;z).$$

Condition 4.1(iv) plays a fundamental role in the theory of relation algebras. It is useful to notice in this connection the following

THEOREM 4.2. *In Definition* 4.1 *condition* (iv) *can be replaced by the following one:*

(iv') *Given any element* $a \in A$, *the functions* f *and* g *defined by the formulas*

$$f(x) = a;x \text{ and } g(x) = a^\smile;x \text{ for every } x \in A$$

are conjugate, and so are the functions f' *and* g' *defined by the formulas*

$$f'(x) = x;a \text{ and } g'(x) = x;a^\smile \text{ for every } x \in A.$$

Proof. By 1.11 and 4.1.

We shall not develop here either the arithmetic of relation algebras or the proper algebraic theory of these algebras (the study of isomorphisms, homomorphisms, subalgebras, etc.)—except insofar as it is relevant for the main purposes of our discussion. Some arithmetical consequences of 4.1 are stated in the next theorem.

THEOREM 4.3. *For any relation algebra*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', ^\smile\rangle$$

---

[15] The axiom system 4.1(i)-(iv) is equivalent to the one given in Jónsson-Tarski [2]. For the proof of the equivalence of the two systems see Chin-Tarski [1], Theorem 2.2; for the relation of these systems to the axiom system in Tarski [2] see Chin-Tarski [1], footnote 10.

we have:

(i) *If* $x \in A$, *I is an arbitrary set, and the elements* $y_i \in A$ *with* $i \in I$ *are such that* $\sum_{i \in I} y_i \in A$, *then*

$$\sum_{i \in I} (x ; y_i) \in A, \quad \sum_{i \in I} (y_i ; x) \in A,$$
$$x ;(\sum_{i \in I} y_i) = \sum_{i \in I} (x ; y_i), \text{ and } (\sum_{i \in I} y_i) ; x = \sum_{i \in I} (y_i ; x).$$

(ii) *If* $x, x', y, y' \in A$, $x \leq x'$, *and* $y \leq y'$, *then* $x ; y \leq x' ; y'$.

(iii) $x ; 0 = 0 = 0 ; x$ *for every* $x \in A$.

(iv) $x \leq x ; 1$ *and* $x \leq 1 ; x$ *for every* $x \in A$.

(v) $(x + y)^{\smile} = x^{\smile} + y^{\smile}$, $(x \cdot y)^{\smile} = x^{\smile} \cdot y^{\smile}$, *and* $(x ; y)^{\smile} = y^{\smile} ; x^{\smile}$ *for any* $x, y \in A$.

(vi) *If* $x, y \in A$ *and* $x \leq y$, *then* $x^{\smile} \leq y^{\smile}$.

(vii) $x^{\smile\smile} = x$ *and* $x^{-\smile} = x^{\smile -}$ *for every* $x \in A$.

(viii) $(x ; y) \cdot z \leq x ; x^{\smile} ; z$ *for any* $x, y, z \in A$.

(ix) $x \leq x ; x^{\smile} ; x$ *for every* $x \in A$.

(x) $x^{\smile} ; (x ; y)^{-} \leq y^{-}$ *for any* $x, y \in A$.

(xi) $0^{\smile} = 0$, $1^{\smile} = 1$, *and* $(1')^{\smile} = 1'$.

(xii) *If* $x \in A$ *is an atom, then* $x^{\smile}$ *is an atom.*

Proof. The proof of parts (i)-(xi) of this theorem can be found elsewhere[16]; (xii) obviously follows from (vi), (vii), and (xi).

THEOREM 4. 4. *Every relation algebra is a normal Boolean algebra with operators.*

Proof. By 2. 13, 4. 1, and 4. 3 (i) (iii) (v) (xi).

Some further arithmetical notions applying to relation algebras are introduced in the following.

DEFINITION 4. 5. *Let*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', {}^{\smile} \rangle$$

be a relation algebra.

(i) *The element* $(1')^{-}$ *is called the diversity element and is denoted by* $0'$. *An element* $x \in A$ *is referred to as*

(ii) *an equivalence element if* $x ; x \leq x$ *and* $x^{\smile} \leq x$,

[16] See Chin-Tarski [1], §§ 1 and 2.

(iii) *a functional element if* $x^{\smile} ; x \leq 1'$,

(iv) *an ideal element if* $x = 1 ; x ; 1$.

The reasons why we have chosen the terms introduced in 4. 5(ii)(iii)(iv) will become clear in our further discussion (see Theorem 4. 24 and remarks preceding Definition 4. 8).

THEOREM 4. 6. *Given a relation algebra*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', {}^{\smile} \rangle$$

we have

(i) $(0')^{\smile} = 0'$.

(ii) $0' ; 0' ; 0' ; 0' = 0' ; 0'$.

(iii) *For every* $x \in A$ *the following three conditions are equivalent:* x *is an equivalence element,* $x ; x = x = x^{\smile}$, *and* $x^{\smile} ; x = x$.

(iv) *If* $x, y \in A$, x *is a functional element, and* $y \leq x$, *then* y *is a functional element.*

(v) *If* $x, y, z \in A$, *and* x *is a functional element, then*

$$x ; (y \cdot z) = (x ; y) \cdot (x ; z) \text{ and } (y \cdot z) ; x^{\smile} = (y ; x^{\smile}) \cdot (z ; x^{\smile}).$$

(vi) *If* $x \in A$ *and* $x \leq 1'$, *then* x *is both an equivalence element and a functional element.*

(vii) *If* $x, y \in A$, x *is an atom, and* y *is a functional element, then* $x ; y = 0$ *or* $x ; y$ *is an atom.*

(viii) *If* $x, y \in A$ *are ideal elements, then* $x + y$, $x \cdot y$, *and* x^{-} *are ideal elements.*

(ix) *If* $x, y \in A$ *are ideal elements, then* $x ; y = x \cdot y$ *and* $x^{\smile} = x$.

(x) *If* $x, y, z \in A$ *and* x *is an ideal element, then*

$$x \cdot (y ; z) = (x \cdot y) ; (x \cdot z).$$

(xi) 0 *and* 1 *are ideal elements.*

(xii) $1 ; x ; 1$ *is an ideal element for every* $x \in A$.

(xiii) $x \in A$ *is an ideal element if, and only if,* x *and* x^{-} *are equivalence elements.*

Proof. The proof of all parts of this theorem, except (vii), can be found elsewhere.[17] To prove (vii), suppose x is an atom and y is a func-

[17] See Chin-Tarski [1], § 3.

tional element. Consider any element $z \neq 0$ such that $z \leq x ; y$. Then $(x ; y) \cdot z \neq 0$ and hence, by 4.1(iv), $(z ; y^{\smile}) \cdot x \neq 0$. Since x is an atom, this implies that $x \leq z ; y^{\smile}$. Hence, by 4.1(ii)(iii), 4.3(ii), and 4.5(iii),

$$x ; y \leq z ; y^{\smile} ; y \leq z ; 1' = z,$$

and therefore $z = x ; y$. Since this is true for every $z \neq 0$ with $z \leq x ; y$, we obtain the conclusion of (vii) at once.

We shall now establish some facts belonging to the proper algebraic theory of relation algebras.

THEOREM 4.7. (i) *A homomorphic image of a relation algebra is again a relation algebra.*

(ii) *A cardinal product of relation algebras is again a relation algebra.*

Proof. The corresponding theorem for Boolean algebras is well known. Hence and from the form of conditions (ii), (iii), and (iv) of 4.1 we see directly that 4.7(ii) holds; to obtain 4.7(i) we notice in addition that, by 1.15(i)(iii) and 4.2, condition 4.1(iv) can be equivalently replaced by a system of equations.

By an *ideal* in the relation algebra

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, {}^{\smile} \rangle$$

we understand any non-empty set $J \subseteq A$ satisfying the conditions:

(i) if $x, y \in J$, then $x + y \in J$;

(ii) if $x \in J$ and $y \in A$, then $x \cdot y \in J$, $x ; y \in J$, and $y ; x \in J$.[18]

The connections between homomorphisms and ideals in relation algebras prove to be entirely analogous to those in Boolean algebras or arbitrary rings. Moreover, it turns out that the discussion of ideals in relation algebras reduces entirely to the discussion of ideals in Boolean algebras. For a one-to-one correspondence which preserves the inclusion relation can be established between ideals in the relation algebra $\mathfrak{A}$ and those in the Boolean algebra

$$\mathfrak{J} = \langle I, +, 0, \cdot, 1 \rangle$$

where I is the set of all ideal elements $a \in A$. (The system $\mathfrak{J}$ is clearly a Boolean algebra by 4.6(viii)(xi).) In particular, given any $a \in I$, the set of all elements $x \in A$ with $x \leq a$ is a principal ideal in $\mathfrak{A}$ which corresponds

[18] See Chin-Tarski [1], § 3 (remarks following Definition 3.23).

to the principal ideal in $\mathfrak{J}$ consisting of all elements $x \varepsilon I$ with $x \leq a$; if a is not in I, the set of all $x \varepsilon A$ with $x \leq a$ is not an ideal in $\mathfrak{A}$. Without using the notion of an ideal explicitly, we give a few theorems, 4.9-4.14, which are suggested by the above remarks and by the knowledge of analogous results applying to Boolean algebras.

DEFINITION 4.8. *Let*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', ^{\smile} \rangle$$

be a relation algebra. Let a be any element in A, and B be the set of all elements $x \varepsilon A$ such that $x \leq a$. Then the system

$$\langle B, +, 0, \cdot, a, ;, a \cdot 1', ^{\smile} \rangle$$

will be denoted by $\mathfrak{A}(a)$.

THEOREM 4.9. *If*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', ^{\smile} \rangle$$

is a relation algebra and $a \varepsilon A$ is an ideal element, then $\mathfrak{A}(a)$ is a relation algebra and the function ϕ on A defined by the formula

$$\phi(x) = a \cdot x \textit{ for every } x \varepsilon A$$

maps $\mathfrak{A}$ homomorphically onto $\mathfrak{A}(a)$.

Proof. It is well known that ϕ maps homomorphically the Boolean algebra $\langle A, +, 0, \cdot, 1 \rangle$ onto $\langle B, +, 0, \cdot, a \rangle$. Moreover, by 4.6(x) we have

$$\phi(x \,;\, y) = \phi(x) \,;\, \phi(y) \text{ for any } x, y \varepsilon A,$$

while, by 4.3(v) and 4.6(ix),

$$\phi(x^{\smile}) = [\phi(x)]^{\smile} \text{ for every } x \varepsilon A.$$

Since, in addition, $\phi(1') = a \cdot 1'$, we conclude by 4.8 that ϕ maps $\mathfrak{A}$ homomorphically onto $\mathfrak{A}(a)$. Hence, by 4.7(i), $\mathfrak{A}(a)$ is a relation algebra.

Notice that a necessary and sufficient condition for $A(a)$ to be a relation algebra is that a be an equivalence element (and not necessarily an ideal element).

THEOREM 4.10. *Let*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', ^{\smile} \rangle$$

be a relation algebra in which $0 \neq 1$. *Then the following conditions are equivalent:*

(i) $\mathfrak{A}$ *is simple.*

(ii) $\mathfrak{A}$ *has no ideal elements different from* 0 *and* 1.

(iii) *For every* $x \in A$, $x \neq 0$ *implies* $1 ; x ; 1 = 1$.

(iv) *For any* $x, y \in A$, $x ; 1 ; y = 0$ *implies that* $x = 0$ *or* $y = 0$.[19]

Proof. By 4.9, (i) implies (ii). By 4.3(iv) and 4.6(xii), (ii) implies (iii). Now assume that (iii) holds. Let ϕ be a homomorphism mapping $\mathfrak{A}$ onto another algebra

$$\mathfrak{B} = \langle B, +, 0, \cdot, 1, ;, 1', \smile \rangle$$

and suppose B contains at least two different elements. By 4.7(i), $\mathfrak{B}$ is also a relation algebra. If $x, y \in A$ and $\phi(x) = \phi(y)$, and if we put

$$z = x \cdot y^{-} + y \cdot x^{-},$$

then $\phi(z) = 0$, and hence, by 4.3(iii), $\phi(1 ; z ; 1) = 0$. Hence $\phi(1 ; z ; 1) \neq 1 = \phi(1)$, and therefore $1 ; z ; 1 \neq 1$. It follows by (iii) that $z = 0$ whence $x = y$. Thus, ϕ is an isomorphism. Consequently, (iii) implies (i), and conditions (i)-(iii) are equivalent.

Suppose now again that (iii) holds, and consider any elements $x, y \in A$ such that $x ; 1 ; y = 0$. By 4.3(iii) we obtain $1 ; x ; 1 ; y = 0$. Hence, by (iii), if $x \neq 0$, then $1 ; y = 0$ and therefore, by 4.3(iv), $y = 0$. Thus (iii) implies (iv). Finally, assume (iv) to hold. From 4.3(x), with x and y replaced by $1 ; x$ and 1, respectively we obtain $(1 ; x)^{\smile} ; (1 ; x ; 1)^{-} \leq 0$, and therefore, by 4.3(v)(xi), $x^{\smile} ; 1 ; (1 ; x ; 1)^{-} = 0$. Hence, by applying (iv) and with the help of 4.3(vii)(xi), we conclude that, if $x \neq 0$, then $(1 ; x ; 1)^{-} = 0$ and $1 ; x ; 1 = 1$. Thus, (iv) implies (iii), and the proof is complete.

THEOREM 4.11. *Every relation algebra which is a subalgebra of a simple relation algebra is itself simple.*

Proof. By 4.10(i)(iii).

THEOREM 4.12. *Let*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', \smile \rangle$$

be a relation algebra. Let I be an arbitrary set, and, for every $i \in I$, *let*

$$\mathfrak{B}_i = \langle B_i, +_i, 0_i, \cdot_i, 1_i, ;_i, 1'_i, \smile_i \rangle$$

---

[19] This is a joint result of J. C. C. McKinsey and A. Tarski; see Jónsson-Tarski [2].

be a relation algebra. In order that $\mathfrak{A}$ be isomorphic to the cardinal product of the algebras $\mathfrak{B}_i$ it is necessary and sufficient that there exist elements $a_i \varepsilon A$ satisfying the following conditions:

(i) $$\sum_{i \varepsilon I} a_i = 1.$$

(ii) *If $i, j \varepsilon I$ and $i \neq j$, then $a_i \cdot a_j = 0$.*

(iii) $\sum_{i \varepsilon I} (x_i \cdot a_i) \varepsilon A$ *for any elements $x_i \varepsilon A$ correlated with $i \varepsilon I$.*

(iv) *a_i is an ideal element for every $i \varepsilon I$.*

(v) *$\mathfrak{A}(a_i) \cong \mathfrak{B}_i$ for every $i \varepsilon I$.*

(*Condition* (iii) *is automatically satisfied in case the set I is finite.*)

Proof. Let

$$\mathfrak{B} = \langle B, +, 0, \cdot, 1, ;, 1', \smile \rangle$$

be the cardinal product of all the algebras $\mathfrak{B}_i$ with $i \varepsilon I$. B is then the set of all functions ϕ on I such that $\phi(i) \varepsilon B_i$ for every $i \varepsilon I$. Given $i \varepsilon I$, let b_i be the unique function $\phi \varepsilon B$ such that

$$\phi(i) = 1_i, \text{ and } \phi(j) = 0_j \textit{ for every } j \varepsilon I, j \neq i.$$

As is easily seen, using 4. 3 (iii) (iv) and 4. 5 (iv), the elements $b_i \varepsilon B$ thus defined satisfy conditions (i)-(v) of our theorem (with 'A' and 'a_i' changed to 'B' and 'b_i'). Hence, if $\mathfrak{A}$ is isomorphic to $\mathfrak{B}$, there are also elements $a_i \varepsilon A$ satisfying the same conditions.

Assume now, conversely, that there are elements $a_i \varepsilon A$ satisfying conditions (i)-(v). For any given $x \varepsilon A$ let $\psi(x)$ be the only function ϕ on I such that

$$\phi(i) = a_i \cdot x \text{ for every } i \varepsilon I.$$

Then, by (i)-(iii), ψ maps the set A in one-to-one way onto the set of all functions ϕ on I such that $\phi(i) \leq a_i$ for every $i \varepsilon I$. Hence, with the help of (iv), 4. 8, and 4. 9, we conclude that ψ maps $\mathfrak{A}$ isomorphically onto the direct product of the algebras $\mathfrak{A}(a_i)$ with $i \varepsilon I$. Consequently, by (v), $\mathfrak{A}$ is isomorphic to the direct product $\mathfrak{B}$ of the algebras $\mathfrak{B}_i$ with $i \varepsilon I$. Thus, our theorem holds in both directions.

THEOREM 4. 13. *Let*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', \smile \rangle$$

as well as $\mathfrak{B}$ and $\mathfrak{C}$ be relation algebras. In order that $\mathfrak{A} \cong \mathfrak{B} \times \mathfrak{C}$ it is necessary and sufficient that there be an ideal element $a \varepsilon A$ such that

$$\mathfrak{A}(a) \cong \mathfrak{B} \text{ and } \mathfrak{A}(a^-) \cong \mathfrak{C}.$$

Proof. By 4.6(viii) and 4.12.

THEOREM 4.14. *A relation algebra is simple if, and only if, it is indecomposable.*

Proof. By 4.8, 4.10, and 4.13 a relation algebra which is not simple cannot be indecomposable. The converse is known to hold for every algebra.

THEOREM 4.15. *For every relation algebra $\mathfrak{A}$ there exist simple relation algebras $\mathfrak{B}_i$ correlated with elements i of a set I such that $\mathfrak{A}$ is isomorphic to a subalgebra of the cardinal product of the algebras $\mathfrak{B}_i$ and that, for every $i \in I$, $\mathfrak{B}_i$ is a homomorphic image of $\mathfrak{A}$.*

Proof. If we replace in 4.15 "relation algebras" by "algebras" and "simple" by "indecomposable," we obtain a statement which holds for arbitrary algebras and is a direct consequence of a result known from the literature.[20] Hence, by restricting ourselves to relation algebras and applying 4.7(i) and 4.14, we obtain 4.15 at once.

It may be mentioned that relation algebras and specifically simple relation algebras are closely related to cylindric algebras discussed in the preceding section. In fact, if

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', ^{\smile} \rangle$$

is a relation algebra, and if we put

$$C_0(x) = 1 ; x \text{ and } C_1(x) = x ; 1 \text{ for every } x \in A,$$

then

$$\mathfrak{A}' = \langle A, +, 0, \cdot, 1, C_0, C_1 \rangle$$

proves to be a generalized cylindric algebra. Moreover, if $\mathfrak{A}$ is simple, then $\mathfrak{A}'$ is also simple, and hence $\mathfrak{A}'$ is a cylindric algebra in the sense of 3.15 (compare remarks following 3.18).

Besides simple algebras, a more special class of relation algebras—in fact, that of integral algebras defined below in 4.16—will be involved in a part of our further discussion (see Theorems 5.10-5.12 in the following section). An even more important role will be played by still another class of relation algebras—in fact, by algebras in which every atom is a functional element; we do not introduce any special term to denote such algebras. Some general properties of these two classes of relation algebras will be established in the following theorems.

[20] See Birkhoff [2].

DEFINITION 4. 16. *A relation algebra*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', {}^{\smile} \rangle$$

is said to be an integral relation algebra (or a relation algebra without zero divisors) if $0 \neq 1$ *and if, for any* $x, y \varepsilon A$, *the formula* $x ; y = 0$ *implies that* $x = 0$ *or* $y = 0$.

THEOREM 4. 17. *For every relation algebra*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', {}^{\smile} \rangle$$

in which $0 \neq 1$ *the following conditions are equivalent:*

(i) $\mathfrak{A}$ *is an integral relation algebra.*

(ii) *For every* $x \varepsilon A$, $x \neq 0$ *implies* $x ; 1 = 1$.

(iii) *Every functional element* $x \varepsilon A$ *such that* $x \neq 0$ *is an atom.*

(iv) $1'$ *is an atom.*

Proof. Assume (i) to hold. By 4. 3 (x) with $y = 1$ we have $x^{\smile} ; (x ; 1)^{-} = 0$ for every $x \varepsilon A$. Hence, by (i) and 4. 16, $x^{\smile} = 0$ or $x ; 1 = 1$. If $x \neq 0$, then, by 4. 3 (vii) (xi), $x^{\smile} \neq 0$ and therefore $x ; 1 = 1$. Thus, (i) implies (ii).

Suppose that (ii) holds. By 4. 3 (viii), with $y = 1$ and $z = 1'$, and 4. 1 (iii), we obtain for every $y \varepsilon A$:

(1) If $y \neq 0$, then $1' = (y ; 1) \cdot 1' \leq y ; y^{\smile} ; 1' = y ; y^{\smile}$.

Let now $x \neq 0$ be a functional element and let y be any element such that $y \leqq x$ and $y \neq 0$. By 4. 1 (iii). 4. 3 (ii) (vi), 4. 5 (iv), and (1) we have

$$x = 1' ; x \leq y ; y^{\smile} ; x \leq y ; x^{\smile} ; x \leq y ; 1' = y.$$

Hence $y = x$. Thus, x is an atom, and (ii) implies (iii).

By 4. 6 (vi), $1'$ is a functional element. Also, $1' \neq 0$ since otherwise we should have, by 4. 1 (iii) and 4. 3 (iii).

$$1 = 1 ; 1' = 1 ; 0 = 0,$$

contrary to the hypothesis of the theorem. Hence, (iii) implies (iv).

Now assume that (iv) holds, and consider any elements $x, y \varepsilon A$ for which $x ; y = 0$. By (iv), either $(y ; 1) \cdot 1' = 1'$ or $(y ; 1) \cdot 1' = 0$. In the first case we have $1' \leqq y ; 1$, and hence, by 4. 1 (ii) (iii) and 4. 3 (ii) (iii),

$$x = x ; 1' \leq x ; y ; 1 = 0 ; 1 = 0.$$

In the second case we obtain, by 4. 1 (iv), $(y^{\smile} ; 1') \cdot 1 = 0$, and hence, by

4.1(iii) and 4.3(vii)(xi), $y = 0$. Consequently, by 4.16, (iv) implies (i). Thus, conditions (i)-(iv) are equivalent.

From 4.3(v)(vii)(xi) it is easily seen that formula $x\, ;1 = 1$ in 4.17(ii) can be replaced by $1\, ;x = 1$.

THEOREM 4.18. (i) *Every relation algebra which is a subalgebra of an integral relation algebra is itself integral.*

(ii) *Every integral relation algebra is simple.*

Proof. (i) immediately follows from 4.16. To obtain (ii) we notice that, by 4.3(ii) (iv) and 4.16, condition 4.17(i) directly implies 4.10(iv). (Similarly, 4.17(ii) directly implies 4.10(iii).)

THEOREM 4.19. *For every relation algebra*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', ^{\smile} \rangle$$

the following conditions are equivalent:

(i) *Every atom $x \in A$ is a functional element.*

(ii) *If $x, y \in A$ are atoms, then $x\, ;y = 0$ or $x\, ;y$ is an atom.*

(iii) *If $x \in A$ is an atom, then $x^{\smile}\, ;x$ is an atom.*

Proof. Observe that, for every $x \in A$, $(x^{\smile}\, ;x) \cdot 1' = 0$ implies $x = 0$ by 4.1(iii)(iv). Hence

$$(1) \qquad (x^{\smile}\, ;x) \cdot 1' \neq 0 \text{ whenever } x \neq 0.$$

By 4.6(vii), condition (i) implies (ii). Assume now (ii) to hold. If $x \neq 0$ is an atom, then $x^{\smile}$ is an atom by 4.3(xii), and therefore $x^{\smile}; x$ is an atom by (ii) and (1). Thus, (ii) implies (iii). From (1) and 4.5(iii) we also see that (iii) implies (i), and the proof is complete.

THEOREM 4.20. *For every atomistic relation algebra*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', ^{\smile} \rangle$$

in which $0 \neq 1$ the following two conditions are equivalent:

(i) *$\mathfrak{A}$ is integral and every atom $x \in A$ is a functional element.*

(ii) *If $x, y \in A$ are atoms, then $x\, ;y$ is also an atom.*

Proof. By 4.16 and 4.19(i)(ii), condition (i) implies (ii) for every relation algebra $\mathfrak{A}$ (whether atomistic or not). Again by 4.19(i)(ii), condition (ii) implies that every atom in $\mathfrak{A}$ is a functional element, and since $\mathfrak{A}$ is atomistic, this condition also implies by 4.3(ii) and 4.16 that $\mathfrak{A}$ is integral. Thus conditions (i) and (ii) are equivalent.

We now turn to the main subject of this section—representation theorems for relation algebras. We begin with the extension theorem:

THEOREM 4.21. *For every relation algebra* $\mathfrak{B}$ *there is a complete and atomistic relation algebra* $\mathfrak{A}$ *which is a perfect extension of* $\mathfrak{B}$. *If* $\mathfrak{B}$ *is simple, or integral, the same applies to* $\mathfrak{A}$.

Proof. By 2.15 and 4.4 there is a complete and atomistic Boolean algebra with operators $\mathfrak{A}$ which is a perfect extension of $\mathfrak{B}$. From 2.11(i), 2.18, and 4.1 it follows that $\mathfrak{A}$ is a relation algebra. The second part of the theorem results from 2.11(ii), 4.10(i) (iii), and 4.17(i) (ii).

THEOREM 4.22. *Every relation algebra*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', \smile \rangle$$

is isomorphic to a relation algebra

$$\mathfrak{A}' = \langle \mathsf{A}, \cup, \wedge, \odot, V, |, I, ^{-1} \rangle$$

where A *is a family of binary relations* $R \subseteq V$, V *is an equivalence relation, and* I *is the identity function on the field of* V, *while* $\cup$, $\wedge$, $|$, *and* $^{-1}$ *have their usual (set-theoretical) meaning.*

Proof. By 4.21 no loss of generality arises if we restrict ourselves to the case when $\mathfrak{A}$ is atomistic. Let, in this case, U be the set of all atoms of $\mathfrak{A}$. We define a function F on A by means of the formula

(1) $$F(x) = \underset{\langle a,b\rangle}{\mathrm{E}}\,[a, b \in U \text{ and } a \leq x \,;\, b] \text{ for every } x \in A.$$

Furthermore, we put

(2) $$\mathsf{A} = F^*(A), \qquad V = F(1), \qquad \text{and } I = F(1').$$

Assume that $F(x) = F(y)$ for any given $x, y \in A$. $\mathfrak{A}$ being atomistic, this assumption implies by (1) that

(3) $$x \,;\, b = y \,;\, b \text{ for every } b \in U.$$

By 4.1(iii) and 4.3(i), again in view of the atomistic character of $\mathfrak{A}$, we have

(4) $$x = x \,;\, 1' = x \,;\, (\sum_{1' \geq b \in U} b) = \sum_{1' \geq b \in U} (x \,;\, b),$$

and similarly

(5) $$y = \sum_{1' \geq b \in U} (y \,;\, b).$$

Formulas (3)-(5) imply that $x = y$. Thus, by (2),

(6) F maps A onto A in one-to-one way.

By (1) and 4. 3(i) (iii) we have

(7) $$F(x+y) = F(x) \cup F(y) \text{ for any } x, y \,\varepsilon\, A$$

and

(8) $$F(0) = \Lambda.$$

Consider any $x, y \,\varepsilon\, A$. If $\langle a, b\rangle \,\varepsilon\, F(x\,;y)$, then, by (1), $a \leq x\,;(y\,;b)$ and $a, b \,\varepsilon\, U$. Hence $\mathfrak{A}$ being atomistic, we conclude by 4. 3(i) that there is a $c \,\varepsilon\, U$ such that $a \leq x\,;c$ and $c \leq y\,;b$. Therefore, by (1), $\langle a, c\rangle \,\varepsilon\, F(x)$, $\langle c, b\rangle \,\varepsilon\, F(y)$, and consequently $\langle a, b\rangle \,\varepsilon\, F(x) \mid F(y)$. In a similar way, using (1) and 4. 3(ii), we show that the latter formula implies $\langle a, b\rangle \,\varepsilon\, F(x\,;y)$. Thus,

(9) $$F(x\,;y) = F(x) \mid F(y) \text{ for any } x, y \,\varepsilon\, A.$$

Suppose $x \,\varepsilon\, A$. If $\langle a, b\rangle \,\varepsilon\, [F(x)]^{-1}$, i. e., $\langle b, a\rangle \,\varepsilon\, F(x)$, then by (1), $b \leq x\,;a$ and $a, b \,\varepsilon\, U$. Therefore $(x\,;a)\cdot b \neq 0$ whence $(x^{\cup};b)\cdot a \neq 0$ by 4. 1(iv). Consequently, $a \leq x^{\cup};b$ and $\langle a, b\rangle \,\varepsilon\, F(x^{\cup})$. Thus

$$[F(x)]^{-1} \subseteq F(x^{\cup}) \text{ for every } x \,\varepsilon\, A.$$

Hence the inclusion in the opposite direction can easily be derived with the help of 4. 3(vii), so that finally

(10) $$F(x^{\cup}) = [F(x)]^{-1} \text{ for every } x \,\varepsilon\, A.$$

In view of (6), F has the inverse function F^{-1} which maps A onto A. If we now define the binary operation $\odot$ on elements of A by putting

$$R \odot S = F(F^{-1}(R)\cdot F^{-1}(S)) \text{ for } R, S \,\varepsilon\, \mathsf{A},$$

we clearly have

(11) $$F(x\cdot y) = F(x) \odot F(y) \text{ for any } x, y \,\varepsilon\, A.$$

By (2) and (6)-(11), the functions F maps $\mathfrak{A}$ isomorphically onto

$$\mathfrak{A}' = \langle \mathsf{A}, \cup, \Lambda, \odot, V, \mid, I, {}^{-1}\rangle.$$

Hence $\mathfrak{A}'$ is a relation algebra. By 4. 3(iv)(xi) we have $V \mid V = V = V^{-1}$; thus, V is an equivalence relation. By (1), (2), and 4. 3(iv), the field of V is U. Finally, from (1), (2), and 4. 1(iii) we see that I is the set of all couples $\langle a, a\rangle$ with $a \,\varepsilon\, U$—or, in other words, the identity function on U to U. The proof has thus been completed.

It should be emphasized that, in general, the operation $\odot$ in the algebra $\mathfrak{A}'$ of Theorem 4.22 is not set-theoretical multiplication; i. e., R and S being two relations in A, $R \odot S$ does not necessarily coincide with the intersection $R \cap S$ of R and S. However, the meaning of $\odot$ is unambiguously determined by the fact that, according to 4.1, $\langle \mathsf{A}, \cup, \Lambda, \odot, V\rangle$ is a Boolean algebra; $R \odot S$ is the largest relation in A included in both R and S. As was pointed out at the beginning of this section, the representation theorem 4.22 cannot be improved in the sense that the operation $\odot$ in $\mathfrak{A}'$ cannot be assumed to coincide with set-theoretical multiplication $\cap$. In the remaining part of this section we shall be concerned with those algebras $\mathfrak{A}'$ in which $\odot$ coincides with $\cap$, and we shall prove a few special representation theorems in which such algebras are involved.

DEFINITION 4.23. *A relation algebra*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', {}^\smile\rangle$$

is called a proper relation algebra if A is a family of binary relations, $0 = \Lambda$, and if $R + S = R \cup S$, $R \cdot S = R \cap S$, $R\,;S = R \mid S$, and $R^\smile = R^{-1}$ for any $R, S \varepsilon \mathsf{A}$. $\mathfrak{A}$ is called a proper relation algebra on a set U if, in addition, U is a non-empty set, $1 = U^2$, and $1'$ is the identity function on U.

When referring to proper relation algebras, we shall use the symbol V instead of 1, and the symbol I instead of 1'.

THEOREM 4.24. *Let*

$$\mathfrak{A} = \langle \mathsf{A}, \cup, \Lambda, \cap, V, |, I, {}^{-1}\rangle$$

be a proper relation algebra.

(i) *For a relation $R \varepsilon \mathsf{A}$ to be an equivalence element in $\mathfrak{A}$ it is necessary and sufficient that R be an equivalence relation. In particular, V and I are equivalence relations having the same field.*

(ii) *If I is an identity function, then, for a relation $R \varepsilon \mathsf{A}$ to be a functional element in $\mathfrak{A}$, it is necessary and sufficient that R be a function (a many-to-one relation).*

Proof. The first part of (i) obviously follows from 4.5 (ii) and 4.23. Hence, by 4.6 (vi) (xi) (xiii), V and I are equivalence relations. Since, by 4.1 (i) (iii), $I \subseteq V \subseteq I \mid V$, we see that V and I have the same field. Finally, (ii) is an obvious consequence of 4.5 (iii) and 4.23.

In connection with 4.24 it may be noticed that the notion of an ideal element also assumes a rather simple meaning when applied to proper relation

algebras. In fact, $\mathfrak{A}$ being a proper relation algebra with the universal relation V, let U be the field of V. Since, by 4.24(i), V is an equivalence relation, there exists a partition of U under V, i. e., a family K of non-empty mutually exclusive sets X such that

$$U = \bigcup_{X \varepsilon \mathsf{K}} X \text{ and } V = \bigcup_{X \varepsilon \mathsf{K}} X^2.$$

It is now easily seen that a relation $R \varepsilon \mathsf{A}$ is an ideal element in the sense of 4.5(iv) if, and only if, for some family $\mathsf{L} \subseteq \mathsf{K}$,

$$R = \bigcup_{X \varepsilon \mathsf{L}} X^2.$$

THEOREM 4.25. *For*

$$\mathfrak{A} = \langle \mathsf{A}, \cup, \Lambda, \cap, V, |, I, {}^{-1} \rangle$$

to be a proper relation algebra, it is necessary and sufficient that the following conditions be satisfied:

(i) A *is a set-field whose elements are binary relations and* V *is the universal set of* A.

(ii) $R \mid S \varepsilon \mathsf{A}$ *and* $R^{-1} \varepsilon \mathsf{A}$ *for any* $R, S \varepsilon \mathsf{A}$.

(iii) $I \varepsilon \mathsf{A}$, *and* $R \mid I = R = I \mid R$ *for every* $R \varepsilon \mathsf{A}$.

Proof. Obvious, by 4.1 and 4.23.

THEOREM 4.26. *Let* A *be the family of all binary relations* R *with* $R \subseteq V$, *and let*

$$\mathfrak{A} = \langle \mathsf{A}, \cup, \Lambda, \cap, V, |, I, {}^{-1} \rangle.$$

(i) *For* $\mathfrak{A}$ *to be a proper relation algebra, it is necessary and sufficient that* V *be an equivalence relation and* I *be the identity function on the field of* V.

(ii) *For* $\mathfrak{A}$ *to be a simple proper relation algebra it is necessary and sufficient that, for some non-empty set* U, $V = U^2$ *and* I *be the identity function on* U *(in other words, that* $\mathfrak{A}$ *be a proper relation algebra on a set* U*).*

Proof. If $\mathfrak{A}$ is a proper relation algebra, then V is an equivalence relation by 4.24(i). J being the identity function on the field of V, we clearly have $J \mid I = I$ and also, by 4.25(iii), $J \mid I = J$, so that $I = J$. If, conversely, V is an equivalence relation and I is the identity function on the field of V, we easily see (e. g., by 4.25) that $\mathfrak{A}$ is a proper relation algebra.

Assume that $\mathfrak{A}$ is a simple proper relation algebra. By (i), V is an

equivalence relation and, U being the field of V, I is the identity function on U. Since $V \neq \Lambda$, the set U is not empty. For any given $a \varepsilon U$ we denote by $V(a)$ the set of all couples $\langle x, y\rangle \varepsilon V$ such that $\langle a, x\rangle \varepsilon V$ and $\langle a, y\rangle \varepsilon V$. We easily see that

$$V \mid V(a) \mid V = V(a).$$

Since $V(a) \neq \Lambda$, we conclude by 4.10(i) (iii) that $V(a) = V$. Hence, if $x, y \varepsilon U$, then $\langle x, x\rangle \varepsilon V(a)$, $\langle y, y\rangle \varepsilon V(a)$, and therefore $\langle x, y\rangle \varepsilon V(a) = V$; consequently, $V = U^2$. Thus, the conditions stated in (ii) are necessary for $\mathfrak{A}$ to be simple. If, conversely, $V = U^2$ and I is the identity function on U, then, by (i), $\mathfrak{A}$ is a proper relation algebra. Moreover, we easily check that $V \mid R \mid V = V$ for every $R \subseteq V$ such that $R \neq \Lambda$. Hence, by 4.10(i) (iii), $\mathfrak{A}$ is simple. The proof has thus been completed.

It should be pointed out at this place that, in general the relation I in a proper relation algebra $\mathfrak{A}$ is not always an identity function. Also, even in case $\mathfrak{A}$ is simple, the relation V is not necessarily of the form $V = U^2$, and hence $\mathfrak{A}$ is not necessarily a proper relation algebra on a set U. In this connection the following two theorems deserve attention.

THEOREM 4.27. *Every proper relation algebra*

$$\mathfrak{A} = \langle \mathsf{A}, \cup, \Lambda, \cap, V, \mid, I, {}^{-1}\rangle$$

is isomorphic to a proper relation algebra

$$\mathfrak{A}' = \langle \mathsf{A}', \cup, \Lambda, \cap, V', \mid, I', {}^{-1}\rangle$$

where I' is the identity function on the field of V'.

Proof. For every x let $x^{\odot} = I^{*}(\{x\})$ and for every $R \varepsilon A$ let $F(R)$ be the set of all ordered pairs of the form $\langle x^{\odot}, y^{\odot}\rangle$ with $\langle x, y\rangle \varepsilon R$. Moreover let

$$\mathsf{A}' = F^{*}(\mathsf{A}),\ V' = F(V) \text{ and } I' = F(I).$$

It is easy to check that the function F maps $\mathfrak{A}$ isomorphically onto the system

$$\mathfrak{A}' = \langle \mathsf{A}', \cup, \Lambda, \cap, V', \mid, I'\rangle,$$

whence it follows that $\mathfrak{A}'$ is a relation algebra. Finally, I' is clearly an identity function on the field of V'.

THEOREM 4.28. *For a proper relation algebra $\mathfrak{A}$ to be simple it is necessary and sufficient that $\mathfrak{A}$ be isomorphic to a proper relation algebra $\mathfrak{A}'$ on a set U.*

Proof. U being an arbitrary set, we easily see that $U^2 | R | U^2 = U^2$ for every relation $R \subseteq U^2$ such that $R \neq \Lambda$. Hence, by 4.10 and 4.23, every proper relation algebra $\mathfrak{A}'$ on U is simple, and $\mathfrak{A} \cong \mathfrak{A}'$ is a sufficient condition for $\mathfrak{A}$ to be simple. Assume now, conversely, that

$$\mathfrak{A} = \langle \mathsf{A}, \cup, \Lambda, \cap, V, |, I, ^{-1} \rangle$$

is simple. By 4.24(i) we have

(1) V is a non-empty equivalence relation,

and in view of 4.27, we may assume that

(2) I is the identity function on the field of V.

Take a fixed element a in the field of V and put

(3) $$U = V^*(\{a\}).$$

Let B be the set of all relations $R \subseteq V$. It follows from (1), (2), and 4.26(i) that

$$\mathfrak{B} = \langle \mathsf{B}, \cup, \Lambda, \cap, V, |, I, ^{-1} \rangle$$

is a proper relation algebra. (1) and (3) imply that

$$V | U^2 | V = U^2$$

so that, by 4.5(iv), U^2 is an ideal element in $\mathfrak{B}$. Hence, by 4.9 and 4.23, the function F on B defined by the formula

$$F(R) = U^2 \cap R \text{ for any } R \subseteq V$$

maps $\mathfrak{B}$ homomorphically onto the proper relation algebra

$$\mathfrak{B}(U^2) = \langle \mathsf{B}', \cup, \Lambda, \cap, U^2, |, I', ^{-1} \rangle$$

where B' is the set of all relations $R \subseteq U^2$ while

(4) $$I' = U^2 \cap I$$

$\mathfrak{A}$ is clearly a subalgebra of $\mathfrak{B}$, and hence the same function F maps $\mathfrak{A}$ homomorphically onto the subalgebra

(5) $$\mathfrak{A}' = \langle \mathsf{A}', \cup, \Lambda, \cap, U^2, |, I', ^{-1} \rangle$$

of $\mathfrak{B}(U^2)$, where $\mathsf{A}' = F^*(\mathsf{A})$. Moreover, by (1)-(4), the set U is not empty and I' is the identity function on U. Hence, according to 4.23, $\mathfrak{A}'$ is a proper relation algebra on U. Finally, since $\mathfrak{A}$ is simple and since, by (5), A' has at least two different elements ($U^2 \neq \Lambda$), F maps $\mathfrak{A}$ isomorphically onto $\mathfrak{A}'$. Thus the theorem holds in both directions.

THEOREM 4.29. *If*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', {}^{\smile}\rangle$$

is an atomistic relation algebra in which every atom is a functional element, then $\mathfrak{A}$ *is isomorphic to a proper relation algebra.*

Proof. We repeat the proof of 4.22 by defining U, F, A, V, and I, and deriving conditions (6), (7), (8), (11), and (12). In view of 4.25, our task reduces to showing that

$$F(x \cdot y) = F(x) \cap F(y) \text{ for any } x, y \,\varepsilon\, A.$$

This condition is clearly equivalent to

(1) $\qquad (x \cdot y) ; b = (x ; b) \cdot (y ; b)$ for any $x, y \,\varepsilon\, A$ and $b \,\varepsilon\, U$.

Now, if $b \,\varepsilon\, U$, then $b^{\smile} \varepsilon\, U$ by 4.3(xii). Hence, by hypothesis, $b^{\smile}$ is a functional element, and we obtain (1) by applying 4.3(v)(vii) and 4.6(v).

THEOREM 4.30. *For every algebra*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', {}^{\smile}\rangle$$

the following two conditions are equivalent:

(i) $\mathfrak{A}$ *is isomorphic to a proper relation algebra*

$$\mathfrak{A}' = \langle \mathsf{A}, \cup, \Lambda, \cap, V, |, I, {}^{-1}\rangle$$

where A *is the family of all relations* $R \subseteq V$.

(ii) $\mathfrak{A}$ *is a complete, atomistic relation algebra in which every atom* x *satisfies the formula* $x^{\smile} ; 1 ; x \leq 1'$.

Conditions (i) *and* (ii) *remain equivalent if both the relation* V *in* (i) *is assumed to be of the form* $V = U^2$ *for some non-empty set* U, *and the algebra* $\mathfrak{A}$ *in* (ii) *is assumed to be simple.*[21]

Proof. Assume (i) to hold. Clearly, $\mathfrak{A}'$ is a complete atomistic relation algebra. Every atom $R \,\varepsilon\, \mathsf{A}$ obviously consists of a single ordered couple $\langle x, y\rangle \,\varepsilon\, V$; since, by 4.26(i), I is the identity function on the field of V, we easily check that $R^{-1} \,|\, V \,|\, R \subseteq I$. Hence (ii) follows at once.

[21] A different characterization of algebraic systems which are isomorphic to proper relation algebras constituted by all subrelations of a relation $V = U^2$ is given in McKinsey [2]. Theorem 4.30 implies that the algebraic systems discussed by McKinsey are relation algebras in the sense of 4.1, and in fact that they coincide with complete, atomistic, simple relation algebras in which every atom satisfies the formula $x^{\smile} ; 1 ; x \leq 1'$.

Assume now that (ii) holds. We want first to establish some properties of atoms in $\mathfrak{A}$. Let U be the set of all atoms in $\mathfrak{A}$. From (ii) and 4.3(ii)(iv)(vii)(xii) we easily obtain:

(1) if $u \in U$, then $u^{\smile} \in U$, $u^{\smile} ; u \leq 1'$, and $u ; u^{\smile} \leq 1'$.

Hence, by 4.5(iii) and 4.19,

(2) if $u \in U$, then $u^{\smile} ; u \in U$ and $u ; u^{\smile} \in U$,

and

(3) if $v, w \in U$, then $v ; w \in U$ or $v ; w = 0$.

Consider any $u, v \in U$ such that $u ; u^{\smile} = v ; v^{\smile}$ and $u^{\smile} ; u = v^{\smile} ; v$. We then clearly have

$$u ; u^{\smile} ; v = v ; v^{\smile} ; v = v ; u^{\smile} ; u.$$

Hence, by 4.3(iii)(ix), $u^{\smile} ; v \neq 0$ and $v ; u^{\smile} \neq 0$. Thus $(v ; u^{\smile}) \cdot 1 \neq 0$ whence, by 4.1(iv), $(v^{\smile} ; 1) \cdot u^{\smile} \neq 0$. Therefore, by (1), 4.3(ii)(iv), and (ii), $u^{\smile} \leq v^{\smile} ; 1$ and $u^{\smile} ; v \leq v^{\smile} ; 1 ; v \leq 1'$. Since $u^{\smile} ; v \neq 0$, we obtain $(u^{\smile} ; v) \cdot 1' \neq 0$. By applying 4.1(iii)(iv) we conclude that $(u ; 1') \cdot v \neq 0$ and $u \cdot v \neq 0$, so that finally $u = v$. Thus,

(4) if $u, v \in U$, $u ; u^{\smile} = v ; v^{\smile}$, and $u^{\smile} ; u = v^{\smile} ; v$, then $u = v$.

Now consider any $u, v, w \in U$ such that $u \leq v ; w$. By 4.1(iii), 4.3(ii)(v)(vi), and (1), we conclude that

$$u ; u^{\smile} \leq v ; w ; (v ; w)^{\smile} \leq v ; w ; w^{\smile} ; v^{\smile} \leq u ; 1' ; v^{\smile} \leq v ; v^{\smile}.$$

Hence, by (2), $u ; u^{\smile} = v ; v^{\smile}$. Similarly we obtain $u^{\smile} ; u = w^{\smile} ; w$. Since $u \neq 0$, we have

$$v ; w = (v ; w) \cdot (v ; w) \neq 0.$$

By applying 4.1(iv) twice we arrive at

$$(v^{\smile} ; v ; w) \cdot w \neq 0 \text{ and } (w ; w^{\smile}) \cdot (v^{\smile} ; v) \neq 0.$$

Hence. by (2), $w ; w^{\smile} = v^{\smile} ; v$. Thus,

(5) if $u, v, w \in U$ and $u \leq v ; w$, then $u ; u^{\smile} = v ; v^{\smile}$, $u^{\smile} ; u = w^{\smile} ; w$, and $v^{\smile} ; v = w ; w^{\smile}$.

Finally consider any $v, w \in U$ such that $v^{\smile} ; v = w ; w^{\smile}$, and let $u = v ; w$. Since $(v^{\smile} ; v) \cdot (w ; w^{\smile}) \neq 0$, we obtain, by 4.1(iv), $(v ; w ; w^{\smile}) \cdot v \neq 0$. Hence, by 4.3(iii) and (3), $u = v ; w \neq 0$ and $u \in U$. Thus, in view of (5),

(6) if $v, w \in U$, $v^{\cup}; v = w; w^{\cup}$, and $u = v; w$, then $u \in U$, $u; u^{\cup} = v; v^{\cup}$, and $u^{\cup}; u = w^{\cup}; w$.

We now define a function F on A by putting for every $x \in A$.

(7) $$F(x) = \underset{\langle a,b\rangle}{\mathrm{E}}\,[\textit{for some } u \in U, \quad u \leqq x, \quad a = u; u^{\cup}, \textit{ and } b = u^{\cup}; u].$$

We also put

(8) $$F^*(A) = \mathsf{A}, \quad F(1) = V, \quad \text{and} \quad F(1') = I.$$

If $x, y \in A$ and $F(x) = F(y)$, we conclude from (4) and (7) that the formulas $u \leq x$ and $u \leq y$ are equivalent for every $u \in U$. Hence, $\mathfrak{A}$ being atomistic,

(9) $$F(x) = F(y) \text{ implies } x = y \text{ for any } x, y \in A.$$

Furthermore, we obtain from (7) with the help of (4),

(10) $$F(x + y) = F(x) \cup F(y) \text{ and } F(x \cdot y) = F(x) \cap F(y) \quad \text{for any } x, y \in A,$$

(11) $$F(0) = \Lambda.$$

Also, by 4.3(v)(vi)(vii) and (2), we conclude from (7) that

(12) $$F(x^{\cup}) = [F(x)]^{-1} \text{ for every } x \in A.$$

Let now $x, y \in A$, and assume $\langle a, b\rangle \in F(x; y)$. By (7) we have $u \leq x; y$, $a = u; u^{\cup}$, and $b = u^{\cup}; u$ for some $u \in U$. Hence, by (ii) and with the help of 4.3(i), $u \leq v; w$ for some $v, w \in U$ such that $v \leq x$ and $w \leq y$. By (5) we obtain $a = v; v^{\cup}$, $b = w^{\cup}; w$, and $v^{\cup}; v = w; w^{\cup}$. Let $c = v^{\cup}; v = w; w^{\cup}$. Then, by (7), $\langle a, c\rangle \in F(x)$, $\langle c, b\rangle \in F(y)$, and consequently $\langle a, b\rangle \in F(x) \mid F(y)$. If, conversely, $\langle a, b\rangle \in F(x) \mid F(y)$, then by means of an analogous argument, but going in the opposite direction and using (6) instead of (5), we obtain $\langle a, b\rangle \in F(x; y)$. Hence

(13) $$F(x; y) = F(x) \mid F(y) \text{ for any } x, y \in A.$$

From (8)-(13) we see that the function F maps isomorphically the relation algebra $\mathfrak{A}$ onto the algebra

$$\mathfrak{A}' = \langle \mathsf{A}, \cup, \Lambda, \cap, V, \mid, I, {}^{-1}\rangle$$

Hence, by 4.1 and 4.23, $\mathfrak{A}'$ is a proper relation algebra. The elements of A are relations $R \subseteq V$. Consider any relation $R \subseteq V$. By (7) and (8), for every $\langle a, b\rangle \in R$ there is a $u \in U$ such that $a = u; u^{\cup}$ and $b = u^{\cup}; u$. Since, by (ii), $\mathfrak{A}$ is complete, the sum X of all $u \in U$ thus obtained exists, and we

easily see from (7) that $F(x) = R$; hence, $R \varepsilon \mathsf{A}$. Therefore A consists of all relations $R \subseteq V$, and $\mathfrak{A}'$ is an algebra of the kind described in (i).

Thus conditions (i) and (ii) are equivalent. By 4.26 (i) (ii), these conditions remain equivalent if they are modified in the way indicated in the last part of the theorem.

THEOREM 4.31. *For every relation algebra* $\mathfrak{B}$ *the following conditions are equivalent:*

(i) $\mathfrak{B}$ *is isomorphic to a proper relation algebra* $\mathfrak{B}'$.

(ii) $\mathfrak{B}$ *is a subalgebra of a complete atomistic relation algebra*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', {}^{\smile} \rangle$$

in which every atom x *satisfies the formula* $x^{\smile} ; 1 ; x \leq 1'$.

(iii) $\mathfrak{B}$ *is a subalgebra of an atomistic relation algebra* $\mathfrak{A}$ *in which every atom is a functional element.*

Conditions (i)-(iii) *remain equivalent if both the algebra* $\mathfrak{B}'$ *in* (i) *is assumed to be a proper relation algebra on a set* U, *and the algebra* $\mathfrak{A}$ *in* (ii) *and* (iii) *is assumed to be simple.*

Proof. Assume (i) to hold, and put, according to 4.23,

$$\mathfrak{B}' = \langle \mathsf{B}, \cup, \Lambda, \cap, V, |, I, {}^{-1} \rangle.$$

By 4.24(i), V is an equivalence relation. By 4.27, I may be assumed to be the identity function on the field of V. Let A be the family of all relations $R \subseteq V$. Then, by 4.26,

$$\mathfrak{A}' = \langle \mathsf{A}, \cup, \Lambda, \cap, V, |, I, {}^{-1} \rangle$$

is a proper relation algebra, and $\mathfrak{B}'$ is clearly a subalgebra of $\mathfrak{A}'$. Since $\mathfrak{B} \cong \mathfrak{B}'$, we can construct, by applying the familiar "exchange method," an algebra $\mathfrak{A} \cong \mathfrak{A}'$ such that $\mathfrak{B}$ is a subalgebra of $\mathfrak{A}$. By 4.30, $\mathfrak{A}$ is a complete atomistic relation algebra in which every atom x satisfies the formula $x^{\smile} ; 1 ; x \leqq 1'$. Thus, condition (i) implies (ii). By 4.3 (ii) (iv) and 4.5 (iii), condition (ii) clearly implies (iii), while, by 4.29, (iii) implies (i). Hence, conditions (i)-(iii) are equivalent. By means of an analogous argument we can show that these conditions remain equivalent if they are modified in the way indicated in the last part of the theorem; to derive then (i) from (iii), we apply 4.11 and 4.28, in addition to 4.29.

So far Theorem 4.15 has not been involved in our discussion of the representation problem. Nevertheless, some possibilities of applying this

theorem to the representation problem can easily be anticipated. In fact, as a consequence of 4.15, the representation problem for arbitrary relation algebras reduces to that for simple relation algebras. Speaking more specifically, consider any class K of relation algebras, and let L be the class of all simple relation algebras which are homomorphic images of algebras in K. It is easily seen that, if every relation algebra of a certain class has a natural representation, the same applies to all cardinal products and to all subalgebras of algebras of this class. Hence, if we succeed in showing that every algebra of the class L has a natural representation, then, due to 4.15, this result automatically extends to all algebras of the class K. We do not know, however, any interesting applications of Theorem 4.15 in the direction just indicated. On the other hand, it will be seen from our further discussion that in some cases by means of Theorem 4.15 we can obtain additional information regarding relation algebras of which we have been able to show (without the help of this theorem) that they have a natural representation.

THEOREM 4.32. *Let*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', \smile \rangle$$

be a relation algebra in which the element 1 *can be represented in the form*

$$1 = \sum_{k < m} x_k$$

where $x_0, x_1, \cdots, x_{m-1}$ *are functional elements. Then:*

(i) $\mathfrak{A}$ *is isomorphic to a proper relation algebra.*

(ii) *If* $\mathfrak{A}$ *is simple, it is isomorphic to a proper relation algebra* $\mathfrak{A}'$ *on a set* U *which contains at most* m *different elements.*

(iii) *In the general case,* $\mathfrak{A}$ *is isomorphic to a subalgebra of a cardinal product of proper relation algebras* $\mathfrak{A}_i$ *on sets* U_i, *each of these sets containing at most* m *different elements.*

Proof. By 4.21, there is a complete and atomistic relation algebra

$$\mathfrak{B} = \langle B, +, 0, \cdot, ;, 1', \smile \rangle$$

which is a perfect extension of $\mathfrak{A}$. $\mathfrak{A}$ is of course a subalgebra of $\mathfrak{B}$ (cf. 2.14). The formula

$$1 = \sum_{i < m} x_i$$

stated in the hypothesis continues to hold in $\mathfrak{B}$ (since the summation is finite); and, by 4.5(iii), the elements x_i which are functional in $\mathfrak{A}$ are functional

in $\mathfrak{B}$ as well. Therefore, for every atom $a \varepsilon B$ there is a functional element $x_i \varepsilon B$ such that $a \leq x_i$. Hence, by 4.6(iv), every atom in $\mathfrak{B}$ is a functional element. Consequently, by 4.31(i) (iii), the relation algebra $\mathfrak{A}$ is isomorphic to a proper relation algebra, and conclusion (i) has thus been established.

If, in addition, $\mathfrak{A}$ is simple, we see, by (i) and 4.28, that $\mathfrak{A}$ is isomorphic to a proper relation algebra on a set U. From the hypothesis we conclude, by 4.23 and 4.24(ii), that U^2 can be represented as the union of m many-to-one relations. Hence, for every $x \varepsilon U$, there are at most m different elements y such that $\langle x, y\rangle \varepsilon U^2$. Consequently, U has at most m different elements, and conclusion (ii) is seen to hold.

In the general case (i.e., without assuming that $\mathfrak{A}$ is simple), we easily see by 4.5(iii) that, not only the algebra $\mathfrak{A}$ itself, but also every homomorphic image of $\mathfrak{A}$ satisfies the hypothesis of our theorem. Hence, by applying 4.15 and (ii), we obtain conclusion (iii).

THEOREM 4.33. *Let*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', {}^{\smile}\rangle$$

be a relation algebra in which $0' ; 0' \leq 1'$. *Then:*

(i) $\mathfrak{A}$ *is isomorphic to a proper relation algebra.*

(ii) *If* $\mathfrak{A}$ *is simple, it is isomorphic to a proper relation algebra* $\mathfrak{A}'$ *on a set* U *which contains at most two different elements.*

(iii) *In the general case,* $\mathfrak{A}$ *is isomorphic to a subalgebra of a direct product of proper relation algebras* $\mathfrak{A}_i$ *on sets* U_i, *each of these sets containing at most two different elements.*

Proof. By 4.5(i) we obviously have $1 = 1' + 0'$. By 4.6(vi), the element $1'$ is functional; under the hypothesis of our theorem, in view of 4.5(iii) and 4.6(i), the element $0'$ is also functional. Thus, 1 is a sum of two functional elements. By now applying 4.32, we obtain all the conclusions at once.

To formulate conveniently the last two theorems of this section, we introduce the following notation:

DEFINITION 4.34. *A relation algebra*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', {}^{\smile}\rangle$$

is said to be

(i) *of class* 1, *if* $0' ; 0' = 0$,

(ii) *of class* 2, *if* $0' ; 0' = 1'$,

(iii) *of class* 3, *if* $0' ; 0' = 1$.

THEOREM 4.35. *Every simple relation algebra*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', ^{\smile} \rangle$$

satisfies one and only one of the following three conditions:

(i) *$\mathfrak{A}$ is of class 1 and is isomorphic to a proper relation algebra $\mathfrak{A}'$ on a set U which contains just one element;*

(ii) *$\mathfrak{A}$ is of class 2 and is isomorphic to a proper relation algebra $\mathfrak{A}'$ on a set U which contains just two different elements;*

(iii) *$\mathfrak{A}$ is of class 3 and either is not isomorphic to any proper relation algebra or else is isomorphic to a proper relation algebra $\mathfrak{A}'$ on a set U which contains at least three different elements.*

Proof. From 4.5(i), 4.23 and 4.34(i)(ii) we easily conclude:

(1) Every proper relation algebra $\mathfrak{A}'$ on a set U containing just one element, or just two different elements, is of class 1, or of class 2, respectively.

Hence, by 4.33 and 4.34, if the given relation algebra $\mathfrak{A}$ satisfies the formula $0' ; 0' \leq 1'$, then it satisfies one of the conditions (i) or (ii).

If $\mathfrak{A}$ does not satisfy the above formula, we have, by 4.5(i), $(0' ; 0') \cdot 0' \neq 0$ and therefore, by 4.10,

$$1 ; ((0' ; 0') \cdot 0') ; 1 = 1.$$

Hence we obtain, by 4.1(iii), 4.3(i)(ii), 4.5(i), and 4.6(ii),

$$\begin{aligned} 1 &= (0' + 1') ; ((0' ; 0') \cdot 0') ; (0' + 1') \\ &\leq 0' ; 0' ; 0' ; 0' + 0' ; 0' + 0' ; 0' + 0' ; 0' = 0' ; 0'. \end{aligned}$$

Thus, by 4.34(iii), $\mathfrak{A}$ is in this case of class 3. Also, since the formula $0' ; 0' \leq 1'$ fails, it follows from (1) and 4.34 that $\mathfrak{A}$ cannot be isomorphic to any proper relation algebra on a set U containing at most two different elements. Hence, $\mathfrak{A}$ being simple by hypothesis, we conclude with the help of 4.28 that $\mathfrak{A}$ satisfies condition (iii). This completes the proof.

From Theorem 4.35 it is easily seen that, up to isomorphism, there is only one simple relation algebra of class 1 and there are only two simple relation algebras of class 2. One of these two simple algebras of class 2 has exactly four elements (e.g., the relations Λ, I, D, and U^2, where U is a set containing just two elements, I is the identity function on U, and D is the complement of I to U^2); the other has exactly sixteen elements (e.g., all the relations $R \subseteq U^2$ where U is a set containing just two elements).

As regards arbitrary relation algebras, it easily follows from 4.34 that

none of them can be both of class m and class n for $1 \leqq m < n \leqq 3$ unless it is a one-element algebra. On the other hand, there are relation algebras which are not of class n for any $n = 1, 2, 3$. By 4.15 and 4.34, a relation algebra is of class n if, and only if, it is a subalgebra of a cardinal product of simple relation algebras of class n. In particular, relation algebras of class 1 are of a rather trivial nature. They can be characterized by the simple formula $1' = 1$, or else by the condition that every element in such an algebra is an ideal element. It is seen from 4.1 that every Boolean algebra $\langle B, +, 0, \cdot, 1\rangle$ can be completed to a relation algebra $\langle B, +, 0, \cdot, 1, ;, 1', {}^{\smile}\rangle$ of class 1 by putting $x \,;\, y = x \cdot y$ and $x^{\smile} = x$ for any $x, y \in A$, as well as $1' = 1$. For this reason relation algebras of class 1 are also referred to as *Boolean relation algebras.*

Since a cardinal product of relation algebras of class n $(n = 1, 2, 3)$ is again a relation algebra of class n, we infer from 4.15 and 4.35 that every relation algebra $\mathfrak{A}$ can be represented as a subalgebra of the cardinal product of three algebras $\mathfrak{A}_1$, $\mathfrak{A}_2$, and $\mathfrak{A}_3$ of classes 1, 2, and 3, respectively. This result, however, will be essentially improved in the following

THEOREM 4.36. *Every relation algebra*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', {}^{\smile}\rangle$$

is isomorphic to the cardinal product of relation algebras $\mathfrak{A}_1$, $\mathfrak{A}_2$, *and* $\mathfrak{A}_3$ *of classes* 1, 2, *and* 3 *respectively.*

Proof. We put

(1) $$a_3 = 1 \,;((0' ; 0') \cdot 0') \,; 1, \qquad a_2 = (1 \,; 0' ; 1) \cdot a_3^-,$$ $$\text{and } a_1 = a_2^- \cdot a_3^-.$$

Hence obviously

(2) $$1 = a_1 + a_2 + a_3 \quad \text{and} \quad a_1 \cdot a_2 = a_1 \cdot a_3 = a_2 \cdot a_3 = 0.$$

Furthermore, by (1) and 4.6(viii) (xii),

(3) a_1, a_2, and a_3 are ideal elements.

By 4.9 and 4.12, conditions (2) and (3) imply that $\mathfrak{A}(a_1)$, $\mathfrak{A}(a_2)$, and $\mathfrak{A}(a_3)$ are relation algebras such that

(4) $$\mathfrak{A} \cong \mathfrak{A}(a_1) \times \mathfrak{A}(a_2) \times \mathfrak{A}(a_3).$$

From 4.8 or 4.9 we see that $0, a_i, a_i \cdot 1'$, and $a_i \cdot 0'$ are respectively the zero element, the unit element, the identity element, and the diversity element of

the algebras $\mathfrak{A}(a_i)$, $i = 1, 2, 3$. Hence, in view of (4) and 4.34, in order to complete the proof it suffices to show that

$$(a_1 \cdot 0') \; ; \; (a_1 \cdot 0') = 0, \tag{5}$$

$$(a_2 \cdot 0') \; ; \; (a_2 \cdot 0') = a_2 \cdot 1', \tag{6}$$

and

$$(a_3 \cdot 0') \; ; \; (a_3 \cdot 0') = a_3. \tag{7}$$

These three formulas can be established directly by using the laws of the arithmetic of relation algebras stated in 4.1, 4.3, and 4.6. The following indirect proof is, however, somewhat simpler and shorter.

Assume first that $\mathfrak{A}$ is a simple relation algebra. Then, by 4.35, $\mathfrak{A}$ is of class 1, 2, or 3. If $\mathfrak{A}$ is of class 1, we have, by 4.3(iii)(ix)(xi), 4.5(i), and 4.34,

$$0' \leq 0' ; 0' ; 0' = 0 ; 0' = 0$$

so that $0' = 0$. Hence, by (1), 4.3(iii), and 4.34, $a_3 = 0$, $a_2 = 0$, and $a_1 = 1$. If $\mathfrak{A}$ is of class 2, then it follows from (1), 4.3(iii), 4.10(i)(iii), and 4.34 that $a_3 = 0$, $a_2 = 1$, and $a_1 = 0$. Finally, in case $\mathfrak{A}$ is of class 3, we infer from (1) and 4.10(i)(iii) that $a_3 = 1$, $a_2 = 0$, and $a_1 = 0$. In each of these three cases it easily follows from 4.3(iii) and 4.34 that (5)-(7) are satisfied.

It is easily seen that whenever formulas (5)-(7) hold in given relation algebras, they also hold in cardinal products and subalgebras of these algebras. (No essential difficulty arises from the fact that, in view of (1) and 4.5(i), these formulas implicitly involve the operation of complementation, which is not included in the system of fundamental operations of relation algebras.) Consequently, since formula (5)-(7) have been shown to hold in simple relation algebras, we conclude by means of 4.15 that they also hold in arbitrary relation algebras, and the proof is complete.

To illustrate this theorem, consider the case when $\mathfrak{A}$ is a proper relation algebra,

$$\mathfrak{A} = \langle \mathsf{A}, \cup, \Lambda, \cap, V, |, I, ^{-1} \rangle$$

where V is an equivalence relation with the field U, and I is the identity function on U. Let K be the partition of U under V (see remarks following 4.24). Let K_1, K_2, and K_3 be respectively the families of all those sets $X \in \mathsf{K}$ which have exactly one, exactly two, and at least three different elements. Finally, let

$$V_i = \bigcup_{X \in \mathsf{K}_i} X^2 \text{ for } i = 1, 2, 3.$$

The relations V_i prove to be elements of A; in fact, they can be identified with the elements a_i defined in the proof of 4. 35. It is easily seen that they are ideal elements of $\mathfrak{A}$; they are obviously disjoint, and their union is V. Hence

$$\mathfrak{A} \cong \mathfrak{A}(V_1) \times \mathfrak{A}(V_2) \times \mathfrak{A}(V_3).$$

Finally, by means of a simple direct argument, we can show that the algebra $\mathfrak{A}(V_i)$, for $i = 1, 2, 3$, is of the class i in the sense of 4. 34.

From 4. 36 we can derive some further consequences by means of 4. 15 and 4. 34. For instance, it is easily seen that the formula $0' ; 0' \leq 1'$ characterizes those relation algebras $\mathfrak{A}$ for which the algebra $\mathfrak{A}_3$ of 4. 36 has just one element, and which therefore can be represented as cardinal products of an algebra A_1 of class 1 and an algebra $\mathfrak{A}_2$ of class 2; such algebras can also be characterized (up to isomorphism) as subalgebras of cardinal products of simple relation algebras of classes 1 and 2. Similarly, the formula $0' \leq 0' ; 0'$ characterizes those relation algebras which are isomorphic to cardinal products of an algebra $\mathfrak{A}_1$ of class 1 and an algebra $\mathfrak{A}_3$ of class 3. Finally, the formula $1' \leq 0' ; 0'$ is characteristic for those relation algebras which are representable as cardinal products of an algebra $\mathfrak{A}_2$ of class 2 and an algebra $\mathfrak{A}_3$ of class 3.

Section 5.

Relation Algebras and Brandt Groupoids.

In Section 3 we have established fundamental relations between Boolean algebras with operators and complex algebras of arbitrary algebraic systems (algebras in the wider sense). In Section 4 we have studied a special class of Boolean algebras with operators—in fact, the relation algebras. We now want to discuss connections between relation algebras and a special class of algebraic systems—in fact the (generalized) Brandt groupoids; this class includes in particular all the groups as its members.

A Brandt groupoid is an algebraic system formed by a set U of elements, a binary operation $\cdot$, a distinguished subset of U (the set of identity elements), and a unary operation $^{-1}$. Roughly speaking, the main difference between Brandt groupoids and groups consists in the fact that in a Brandt groupoid the set U is not assumed to be closed under multiplication; the domain of the operation $\cdot$ is not, in general, the whole set U^2, but a subset of U^2. In consequence, a Brandt groupoid may contain many identity elements, i. e., many elements u such that $x \cdot u = x = u \cdot x$ whenever x, $x \cdot u$, and $u \cdot x$ are in U. A precise definition of a Brandt groupoid follows:

DEFINITION 5.1. *An algebraic system (algebra in a wider sense)*

$$\mathfrak{U} = \langle U, \cdot, I, {}^{-1} \rangle$$

(where $\cdot$ *is an operation on a subset of* U^2 *to* U, I *is a subset of* U, *and* ${}^{-1}$ *is an operation on* U *to* U*) is called a generalized Brandt groupoid if the following conditions are satisfied:*

(i) *For any elements* $x, y, z \in U$ *such that* $x \cdot y \in U$ *and* $y \cdot z \in U$ *we have* $(x \cdot y) \cdot z \in U$ *and* $(x \cdot y) \cdot z = x \cdot (y \cdot z)$.

(ii) *For any elements* $x, y, z \in U$ *such that* $x \cdot y \in U$ *and* $x \cdot y = x \cdot z$ *we have* $y = z$.

(iii) *For any elements,* $x, y, z \in U$ *such that* $x \cdot z \in U$ *and* $x \cdot z = y \cdot z$ *we have* $x = y$.

(iv) $x \cdot x = x$ *for every* $x \in I$.

(v) $x^{-1} \cdot x \in I$ *and* $x \cdot x^{-1} \in I$ *for every* $x \in U$.

$\mathfrak{U}$ *is called a Brandt groupoid if, in addition, the following condition holds:*

(vi) *For any elements* $x, z \in I$ *there exists an element* $y \in U$ *such that* $x \cdot y \in U$ *and* $y \cdot z \in U$.[22]

In the next three theorems we state without proof certain arithmetic properties of generalized Brandt groupoids which will be used later. These results can be obtained by methods that are essentially known from the literature.[23]

THEOREM 5.2. *Let*

$$\mathfrak{U} = \langle U, \cdot, I, {}^{-1} \rangle$$

be a generalized Brandt groupoid. For any elements $x, y, z \in U$ *we have:*

(i) *If* $y \in I$ *and* $x \cdot y \in U$, *then* $x \cdot y = x$.

(ii) *If* $x \in I$ *and* $x \cdot y \in U$, *then* $x \cdot y = y$.

(iii) $x \cdot (x^{-1} \cdot x) = x = (x \cdot x^{-1}) \cdot x$.

(iv) *If* $x \cdot y \in I$, *then* $x^{-1} = y$.

(v) $(x^{-1})^{-1} = x$.

(vi) *If* $x \cdot y \in U$, *then* $x^{-1} \cdot x = y \cdot y^{-1}$.

(vii) *If* $x \cdot y \in U$, *then* $(x \cdot y)^{-1} = y^{-1} \cdot x^{-1}$.

---

[22] These algebraic systems were first studied by H. Brandt. It is easily seen that axioms (i)-(vi) in Definition 5.1 are equivalent to axioms I-IV of Brandt [1]; when deriving 5.1(i)-(vi) from Brandt's axioms, we let I be the set of all elements $x \in U$ such that $x \cdot x = x$.

[23] Cf. Brandt [1].

THEOREM 5.2. *An algebraic system*

$$\mathfrak{U} = \langle U, \cdot, I, ^{-1} \rangle$$

(where $\cdot$ is a function on a subset of U^2 to U, $I \subseteq U$, and $^{-1}$ is a function on U to U) is a generalized Brandt groupoid if, and only if, the following conditions are satisfied:

(i) *For every element $x \in U$ there is an element $y \in U$ such that $x \cdot y \in U$.*

(ii) *For any elements $x, y, z \in U$, if $x \cdot y$ and $(x \cdot y) \cdot z$ are in U or $y \cdot z$ and $x \cdot (y \cdot z)$ are in U, then all these elements are in U and $(x \cdot y) \cdot z = x \cdot (y \cdot z)$.*

(iii) *For any element $x \in U$ we have $x \in I$ if, and only if, $x \cdot x = x$.*

(iv) *For any elements $x, y, z \in U$ the formulas*

$$x \cdot y = z, \quad x^{-1} \cdot z = y, \text{ and } z \cdot y^{-1} = x$$

are equivalent.

THEOREM 5.4. *For every generalized Brandt groupoid*

$$\mathfrak{U} = \langle U, \cdot, I, ^{-1} \rangle$$

the following three conditions are equivalent:

(i) *$\mathfrak{U}$ is a Brandt groupoid.*

(ii) *For any elements $x, z \in U$ there is an element $y \in U$ such that $x \cdot y \in U$ and $y \cdot z \in U$.*

(iii) *For any elements $x, y \in U$ there are elements $u, v \in U$ such that $u \cdot y \in U$ and $x = u \cdot y \cdot v$.*

According to Definition 3.8, the complex algebra of a generalized Brandt groupoid

$$\mathfrak{U} = \langle U, \cdot, I, ^{-1} \rangle$$

is the algebra

$$\mathfrak{A} = \langle \mathsf{A}, \cup, \Lambda, \cap, U, \cdot^{*}, I^{*}, ^{-1*} \rangle$$

where A is the family of all subsets of U. The operations $\cdot^{*}$, I^{*}, and $^{-1*}$ are understood in the sense of Definition 3.2. Thus, $\cdot^{*}$ is an operation on A^2 to A; for any sets X and Y in A, $X \cdot^{*} Y$ is the set of all elements $z \in U$ such that $z = x \cdot y$ for some $x \in X$ and $y \in Y$. Similarly, for any $X \in \mathsf{A}$, X^{-1*} is the set of all $y \in U$ such that $y = X^{-1}$ for some $x \in X$. Finally, since I is a subset of U, i. e., a unary relation, I^{*} is an operation of rank 0 such that $I^{*}(\Lambda) = I$. This operation I^{*} will be replaced as usual by the set I

itself which will be treated as a distinguished element in A (see remarks in the introduction). Thus we shall speak of

$$\mathfrak{A} = \langle \mathsf{A}, \cup, \Lambda, \cap, U, \cdot^*, I, {}^{-1*} \rangle$$

as the complex algebra of $\mathfrak{U}$.

THEOREM 5. 5. *Let*

$$\mathfrak{U} = \langle U, \cdot, I, {}^{-1} \rangle$$

be a generalized Brandt groupoid, and let

$$\mathfrak{A} = \langle \mathsf{A}, \cup, \Lambda, \cap, U, \cdot^*, I, {}^{-1*} \rangle$$

be the complex algebra of $\mathfrak{U}$. *Then we have:*

(i) $\mathfrak{A}$ *is a relation algebra with at least two different elements* ($\Lambda \neq U$)—*in fact, a complete atomistic relation algebra in which every atom is a functional element.*

(ii) *If* $\mathfrak{U}$ *is a Brandt groupoid, then* $\mathfrak{A}$ *is simple.*

Proof. By 3. 8, $\mathfrak{A}$ is a complete atomistic Boolean algebra with operators. By 5. 3 (ii), for any sets $X, Y, Z \,\varepsilon\, \mathsf{A}$ (i. e., for any subsets X, Y, Z of U) we have

$$(X \cdot^* Y) \cdot^* Z = X \cdot^* (Y \cdot^* Z).$$

From 5. 1 (v) and 5. 2 (i)-(iii) we conclude that

$$I \cdot^* X = X = X \cdot^* I.$$

By 5. 3 (iv) the formulas

$$(X \cdot^* Y) \cap Z = \Lambda,\ (X^{-1*} \cdot^* Z) \cap Y = \Lambda, \text{ and } (Z \cdot^* Y^{-1*}) \cap X = \Lambda$$

are equivalent. Hence, by 4. 1, $\mathfrak{A}$ is a relation algebra. Since $U \neq \Lambda$, $\mathfrak{A}$ has at least two different elements.

Every atom X in $\mathfrak{A}$ is clearly a set of the form $X = \{x\}$ for some $x \,\varepsilon\, U$, and therefore, by 5. 1 (v), it satisfies the formula $X^{-1*} \cdot^* X \subseteq I$. Thus, by 4. 5 (iii), every atom in $\mathfrak{A}$ is a functional element.

If, finally, $\mathfrak{U}$ is a Brandt groupoid, we easily see from 5. 4 (i) (iii) that $U \cdot^* X \cdot^* U = U$ for every $X \,\varepsilon\, \mathsf{A}$ such that $X \neq \Lambda$. Hence, by 4. 10 (i) (iii), $\mathfrak{A}$ is simple.

THEOREM 5. 6. *Let*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', {}^{\smile} \rangle$$

be an atomistic relation algebra, with $0 \neq 1$, *in which every atom is a func-*

tional element. Let U be the set of all atoms of $\mathfrak{A}$, let I be the set of all those atoms u for which $u \leqq 1'$, and let

$$\mathfrak{U} = \langle U, ;, I, {}^{\smile} \rangle.$$

We then have:

(i) *$\mathfrak{U}$ is a generalized Brandt groupoid, and $\mathfrak{A}$ is isomorphic to a subalgebra of the complex algebra of $\mathfrak{U}$.*

(ii) *If $\mathfrak{A}$ is complete, then $\mathfrak{A}$ is isomorphic to the complex algebra of $\mathfrak{U}$.*

(iii) *If $\mathfrak{A}$ is simple, then $\mathfrak{U}$ is a Brandt groupoid.*

Proof. We first want to show that the algebraic system $\mathfrak{U}$ satisfies condition 5. 3(i)-(iv), with $\cdot$ and ${}^{-1}$ replaced by $;$ and ${}^{\smile}$, respectively. (The domain of the operation $;$ is understood to be restricted to the set of those couples $\langle x, y \rangle \in U^2$ for which $x ; y \in U$. An analogous remark applies to the operation ${}^{\smile}$, and from 4. 3(xii) it is seen that the domain of the operation ${}^{\smile}$ thus restricted is U.)

By 4. 3(vii)(xii) and 4. 19(i)(iii) we have $x^{\smile} \in U$ and $x ; x^{\smile} \in U$ for every $x \in U$. Hence 5. 3(i)' holds. If, x, y, z, $x ; y$, and $(x ; y) ; z$ are in U, we have $(x ; y) ; z = x ; (y ; z)$ by 4. 1(ii). Hence, by 4. 3(iii), we conclude that $y ; z \neq 0$, and therefore $y ; z \in U$ by 4. 19(i) (ii). Similarly, under the assumption that x, y, z, $y ; z$, and $x ; (y ; z)$ are in U, we obtain $(x ; y) ; z = x ; (y ; z)$ and $x ; y \in U$. Thus, 5. 3(ii) holds. From 4. 6(iii)(vi) we see that $x ; x = x$ for every $x \in I$. If, conversely, $x ; x = x$ and $x \in U$, we have, by 4. 1(iii), 4. 6(v), and the hypothesis of the theorem, $x = (x ; x) \cdot (x ; 1') = x ; (x \cdot 1')$. Hence, by 4. 3(iii), $x \cdot 1' \neq 0$, therefore $x \cdot 1' = x$ (x being an atom), and consequently $x \in I$. We have thus obtained 5. 3(iii). If, finally, $x, y, z \in U$ and $x ; y = z$, then $(x ; y) \cdot z \neq 0$, hence, by 4. 1(iv), $(x^{\smile} ; z) \cdot y \neq 0$, and therefore $x^{\smile} ; z \neq 0$. Consequently, by 4. 3(xii) and 4. 19, $x^{\smile} ; z$ is an atom, and since y is also an atom, the formula $(x^{\smile} ; z) \cdot y \neq 0$ gives $x^{\smile} ; z = y$. Thus $x ; y = z$ implies $x^{\smile} ; z = y$ for any $x, y, z \in U$. From this implication we easily derive 5. 3(iv) by means of 4. 3(v)(vii).

We now know by 5. 3 that $\mathfrak{U}$ is a generalized Brandt groupoid. We define a function F on A by putting for every $x \in A$

(1) $$F(x) = \mathop{\mathrm{E}}_{u} [u \in U \text{ and } u \leq x].$$

Since $\mathfrak{A}$ is atomistic, this definition clearly implies for any $x, y \in A$:

(2) If $F(x) = F(y)$, then $x = y$ (i. e., F is one-to-one).

(3) $F(x + y) = F(x) \cup F(y)$ and $F(x \cdot y) = F(x) \cap F(y)$.

We also have

(4) $$F(0) = \Lambda, \quad F(1) = U, \text{ and } F(1') = I.$$

By 4. 3 (vi) (vii) (xii) we obtain for every $x \varepsilon A$

(5) $$F(x^{\smile}) = \underset{v}{\mathrm{E}} \, [v \varepsilon U \textit{ and, for some } u, \quad v = u^{\smile} \textit{ and } u \varepsilon F(x)].$$

Finally, we can show that for any $x, y \varepsilon A$

(6) $$F(x \,;\, y) = \underset{w}{\mathrm{E}} \, [w \varepsilon U \textit{ and, for some } u \textit{ and } v, \quad w = u \,;\, v, \quad u \varepsilon F(x), \textit{ and } v \varepsilon F(y)].$$

In fact, if $w \varepsilon F(x \,;\, y)$, then, by (1), $w \varepsilon U$ and $w \leq x \,;\, y$. Hence, with the help of 4. 3 (i) and 4. 19, $w = u \,;\, v$ for some $u, v \varepsilon U$ such that $u \leq x$ and $v \leq y$, and therefore, again by (1), $w = u \,;\, v$ for some $u \varepsilon F(x)$ and $v \varepsilon F(y)$. If, conversely, $w \varepsilon U$ and $w = u \,;\, v$ for some $u \varepsilon F(x)$ and $v \varepsilon F(y)$, then $w \varepsilon F(x \,;\, y)$ by (1) and 4. 3 (ii).

In view of 3. 2 and 3. 8, conditions (1)-(6) show that F maps $\mathfrak{A}$ isomorphically onto a subalgebra $\mathfrak{A}'$ of the complex algebra of $\mathfrak{U}$. If $\mathfrak{A}$ is complete, $\mathfrak{A}'$ clearly coincides with the complex algebra of $\mathfrak{U}$. If $\mathfrak{A}$ is simple, we have, by 4. 10, $x \leq 1 \,;\, y \,;\, 1$ for any $x, y \varepsilon U$, and hence, by 4. 3 (i) (iii) and 4. 19, $x = u \,;\, y \,;\, v$ and $u \,;\, y \varepsilon U$ for some $u, v \varepsilon U$. By 5. 4 this implies that $\mathfrak{U}$ is a Brandt groupoid. The proof has thus been completed.

We now can establish a converse of 5. 5:

THEOREM 5. 7. *Let*

$$\mathfrak{U} = \langle U, R, I, S \rangle$$

be an algebra in the wider sense in which $R \subseteq U^3$, $I \subseteq U$, *and* $S \subseteq U^2$. *Assume that the complex algebra* $\mathfrak{A}$ *of* $\mathfrak{U}$ *is a relation algebra in which every atom is a functional element. Then we have:*

(i) $\mathfrak{U}$ *is a generalized Brandt groupoid.*

(ii) *If* $\mathfrak{A}$ *is simple, then* $\mathfrak{U}$ *is a Brandt groupoid.*

Proof. By 3. 8, 3. 9 and the hypothesis, $\mathfrak{A}$ is a complete atomistic relation algebra, with at least two different elements, in which every atom is a functional element. Hence, by 5. 6, $\mathfrak{A}$ is isomorphic to the complex algebra $\mathfrak{A}'$ of a system $\mathfrak{U}'$ which is a generalized Brandt groupoid; moreover, $\mathfrak{U}'$ is a Brandt groupoid in case $\mathfrak{A}$ is simple. From 3. 2 and 3. 8 we easily see that any two algebras in the wider sense are isomorphic whenever their complex algebras are isomorphic. Hence $\mathfrak{U}$ is isomorphic to $\mathfrak{U}'$; therefore $\mathfrak{U}$ is a generalized Brandt groupoid and, in case $\mathfrak{A}$ is simple, it is a Brandt groupoid.

As is seen from 4. 19, the assumption that every atom in $\mathfrak{A}$ is functional can be replaced in the hypothesis of 5. 7 by the assumption that R is a binary operation (on a subset of U^2 to U), and not only a ternary relation.

THEOREM 5. 8. *For every relation algebra*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', \smile \rangle$$

in which $0 \neq 1$ *the following two conditions are equivalent:*

(i) $\mathfrak{A}$ *is isomorphic to a proper relation algebra,*

(ii) $\mathfrak{A}$ *is isomorphic to a subalgebra of the complex algebra of a generalized Brandt groupoid.*

Also the following two conditions are equivalent:

(i′) $\mathfrak{A}$ *is isomorphic to a proper relation algebra on a set* U,

(ii′) $\mathfrak{A}$ *is isomorphic to a subalgebra of the complex algebra of a Brandt groupoid.*

Proof. Condition (i) implies (ii) by 4. 31 (i) (iii) and 5. 6 (i). Conversely, (ii) implies (i) by 4. 31 (i) (iii) and 5. 5 (i). In an analogous way we show that conditions (i′) and (ii′) are equivalent; we make use of the last part of 4. 31 (concerning a modification of conditions 4. 31 (i)-(iii)), as well as of 5. 5 (ii) and 5. 6 (ii).

The following informal remarks concern an interesting class of generalized Brandt groupoids which can be referred to as having the *unicity property.* Generalized Brandt groupoids

$$\mathfrak{U} = \langle U, \cdot, I, {}^{-1} \rangle$$

of this class are characterized by the following condition:

(i) For any $x, z \in U$ there is at most one element $y \in U$ such that $x \cdot y \in U$ and $y \cdot z \in U$.

Another, equivalent formulation of this condition is

(ii) For any $x, y \in U$, if $x \cdot y \in U$ and $y \cdot x \in U$, then $y = x^{-1}$.

If V is a non-empty equivalence relation, I the set of all couples $\langle x, x\rangle \in V$, and if we define the operations $\cdot$ and ${}^{-1}$ by putting, for any couples $\langle x, y\rangle$, $\langle y, z\rangle \in V$,

$$\langle x, y\rangle \cdot \langle y, z\rangle = \langle x, z\rangle, \quad \langle x, y\rangle^{-1} = \langle y, x\rangle,$$

and by assuming that $\langle x, y\rangle \cdot \langle y', z\rangle$ does not exist in case $y \neq y'$, then, as is easily seen, the system

$$\mathfrak{U}(V) = \langle V, \cdot, I, {}^{-1} \rangle$$

is a generalized Brandt groupoid with the unicity property. Conversely, every generalized Brandt groupoid $\mathfrak{U}$ with the unicity property is isomorphic to a groupoid $\mathfrak{U}(V)$ constructed in the way just described. To obtain a function f mapping $\mathfrak{U}$ on a groupoid $\mathfrak{U}(V)$, we put for every element x in U

$$f(x) = \langle x \cdot x^{-1}, x^{-1} \cdot x \rangle.$$

By comparing these results with 5.6, we easily see that for every relation algebra $\mathfrak{A}$ the following two conditions are equivalent:

(i′) $\mathfrak{A}$ is isomorphic to the proper relation algebra constituted by all subrelations of a non-empty equivalence relation V, and

(ii′) $\mathfrak{A}$ is isomorphic to the complex algebra of a generalized Brandt groupoid $\mathfrak{U}$ with the unicity property.

As a consequence, we obtain by 5.8 that the complex algebra of an arbitrary generalized Brandt groupoid is isomorphic to a subalgebra of the complex algebra of a generalized Brandt groupoid with the unicity property. Either with the help of this result or in a more direct way we conclude that every generalized Brandt groupoid $\mathfrak{U}$ is a homomorphic image of a generalized Brandt groupoid $\mathfrak{U}'$ with the unicity property. In fact, we can take for $\mathfrak{U}'$ the system $\mathfrak{U}(V)$ defined above where V is the set of all couples $\langle x, y \rangle$ such that x, y, and $x \cdot y^{-1}$ are elements in U. The function g defined for every couple $\langle x, y \rangle \varepsilon V$ by the formula

$$g(x, y) = x \cdot y^{-1}$$

maps $\mathfrak{U}(V)$ homomorphically onto $\mathfrak{U}$.

All these remarks remain valid if we replace in them arbitrary generalized Brandt groupoids by Brandt groupoids, and arbitrary equivalence relations by relations of the form $V = W^2$ where W is an arbitrary set.

To conclude this section, we want to give some applications of our results to groups. A group may be considered as a system constituted by a non-empty set U, a binary operation $\cdot$ on U^2 to U, a distinguished element u of U (the unit or identity element), and a unary operation $^{-1}$ on U to U; the postulates which are to be satisfied by these notions are well known. We can, of course replace the element $u \varepsilon U$ by a set $I \subseteq U$ which is assumed to consist of just one element. Groups become then systems of the same type as Brandt groupoids, and we can state the following

THEOREM 5.9. *For*

$$\mathfrak{U} = \langle U, \cdot, I, {}^{-1} \rangle$$

to be a group it is necessary and sufficient that $\mathfrak{U}$ *be a generalized Brandt groupoid in which* $\cdot$ *is an operation on* U^2 *to* U. *The condition remains necessary and sufficient if we omit in it the word "generalized."*

Proof. The theorem easily follows from the definition of a group and that of a (generalized) Brandt groupoid. The only thing which is perhaps not quite obvious is that in a generalized Brandt groupoid in which $\cdot$ is an operation on U^2 (and not only on a subset of U^2) to U the set I consists of just one element. To show this notice that I is non-empty by 5.1(v) and that, for any $x, y \in I$, we have $x \cdot x = x = x \cdot y$ by 5.2(i), and hence $x = y$ by 5.1(ii).

THEOREM 5.10. *The complex algebra of a group is a complete atomistic integral relation algebra in which every atom is a functional element.*[24]

Proof. By 4.16, 5.5(i), and 5.9, with the help of 3.2 and 3.8.

THEOREM 5.11. *Let*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', {}^{\smile}\rangle$$

be an atomistic integral relation algebra in which every atom is a functional element. Let U *be the set of all atoms of* $\mathfrak{A}$, *and let*

$$\mathfrak{U} = \langle U, ;, \{1'\}, {}^{\smile}\rangle.$$

We then have:

(i) $\mathfrak{U}$ *is a group and* $\mathfrak{A}$ *is isomorphic to a subalgebra of the complex algebra of* $\mathfrak{U}$.

(ii) *If* $\mathfrak{A}$ *is complete, it is isomorphic to the complex algebra of* $\mathfrak{U}$.

Proof. By 4.17(i)(iv), $1'$ is an atom. Hence $\{1'\}$ coincides with the set of all atoms $u \leq 1'$, and therefore the system $\mathfrak{U}$ defined in the hypothesis coincides with the system $\mathfrak{U}$ of 5.6. Consequently, by 5.6(i), $\mathfrak{U}$ is a generalized Braandt groupoid. By 4.20, the operation $;$ in $\mathfrak{U}$ is an operation on U^2 to U. Hence, by 5.9, $\mathfrak{U}$ is a group. The remaining conclusions follow directly from 5.6(i) (ii).

THEOREM 5.12. *Let*

$$\mathfrak{U} = \langle U, R, I, S\rangle$$

be an algebra in the wider sense in which $R \subseteq U^3$, $I \subseteq U$, *and* $S \subseteq U^2$. *If the complex algebra* $\mathfrak{A}$ *of* $\mathfrak{U}$ *is an integral relation algebra in which every atom is a functional element, then* $\mathfrak{U}$ *is a group.*

[24] This is a result of J. C. C. McKinsey; see Jónsson-Tarski [2].

Proof. By 5.7, $\mathfrak{U}$ is a generalized Brandt groupoid. From 4.20, with the help of 3.2 and 3.8, we conclude that R is a binary operation on U^2 to U. The conclusion follows by 5.9.

Instead of assuming, in the hypothesis of 5.12, that $\mathfrak{A}$ is integral and that every atom in $\mathfrak{A}$ is a functional element, it suffices to assume that $\mathfrak{U}$ is an algebra, and not only an algebra in the wider sense. Even the (weaker) assumption that R is a binary operation on U^2 to U proves to be sufficient. Compare the analogous remark following 5.7.

The results stated in 5.10 and 5.11 can partly be extended to arbitrary Boolean algebras with operators and to complex algebras of arbitrary algebras (not in the wider sense). A Boolean algebra with operators

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, f_0, f_1, \cdots, f_\xi, \cdots \rangle$$

in which $0 \neq 1$ is called *integral* if, for each of the functions f_ξ with the rank m_ξ and for every sequence $x \in A^{m_\xi}$ the formula $f_\xi(x) = 0$ implies that $x_j = 0$ for some $j < m_\xi$. (In case $\mathfrak{A}$ is a relation algebra with $0 \neq 1$, this condition is automatically satisfied both by $1'$ treated as an operation with the rank 0 and by $^\smile$, and has to be postulated only for $;$.) The condition that $\mathfrak{A}$ is integral is clearly necessary for $\mathfrak{A}$ to be isomorphic to the complex algebra of some algebra. This condition is, however, not sufficient. On the contrary, the following condition—when combined with completeness, atomisticity, and the condition that $0 \neq 1$—is necessary and sufficient for $\mathfrak{A}$ to be isomorphic to the complex algebra of some algebra: for each of the functions f_ξ with the rank m_ξ and for every sequence $x \in U^{m_\xi}$ where U is the set of all atoms in $\mathfrak{A}$ we have $f_\xi(x) \in U$. It is seen from 4.3(xii), 4.17(i)(iv), and 4.20 that the latter condition, when applied to an atomistic relation algebra $\mathfrak{A}$, is equivalent to the one occurring in 5.10 and 5.11, i.e., to the condition that $\mathfrak{A}$ is integral and that every atom in $\mathfrak{A}$ is a functional element.

By 4.18(i) and 5.10, every relation algebra which is a subalgebra of the complex algebra of a group is integral. The question whether, conversely, every integral relation algebra is isomorphic to a subalgebra of the complex algebra of a group is still open. The answer is not known even for those integral algebras which are isomorphic to proper relation algebras.

A REMARK ON BOOLEAN ALGEBRAS WITH OPERATORS.*

By Hugo Ribeiro.

Bjarni Jónsson and Alfred Tarski, in their paper *Boolean algebras with operators* (**73**, 891 and **74**, 127 of this Journal), call a Boolean algebra B a regular subalgebra of a Boolean algebra A if A is complete and atomistic and B is a subalgebra of A for which: i) if I is an arbitrary set and if the elements $x_i \varepsilon B$ with $i \varepsilon I$ are such that $\sum_{i \varepsilon I} x_i = 1$, then there exists a finite subset J of I such that $\sum_{i \varepsilon J} x_i = 1$, ii) if u and v are distinct atoms of A then there exists an element $b \varepsilon B$ such that $u \leqq b$ and $v \cdot b = 0$ (Definition 1.19). The set C of all "closed" elements of A is then defined as the set of all elements $x \varepsilon A$ such that $x = \prod_{x \leq y \varepsilon B} y$ (Definition 1.20); and A^m, B^m, C^m designate the sets of all m-termed sequences, $x = \langle x_0, \cdots, x_{m-1} \rangle$, of elements of A, B, C respectively. Furthermore, a function f on B^m to B is called monotonic if given two sequences $x \leq y \varepsilon B^m$ (that is $x_i \leq y_i \varepsilon B$ for any $i = 0, \cdots, m-1$) we always have $f(x) \leq f(y)$, additive if given any $j < m$ and $x, y \varepsilon B^m$ such that $x_p = y_p$ whenever $j \neq p < m$ we always have $f(x+y) = f(x) + f(y)$ (Definition 1.1); f^+/C^m designates the restriction of the function f^+ to C^m and by the composition $f[g_0, \cdots, g_{m-1}]$ of f, on B^m to B, with $g_0, \cdots, g_{m-1}$, on B^n to B, it is understood the function h on B^n to B such that $h(x) = f(g_0(x), \cdots, g_{m-1}(x))$ whenever $x \varepsilon B^n$. Finally, to any function f on B^m to B an extension, f^+, on A^m to A is defined by $f^+(x) = \sum_{x \geq y \varepsilon C^m} \prod_{y \leq z \varepsilon B^m} f(z)$ for any $x \varepsilon A^m$ (Definition 2.1), and it is shown (as an immediate consequence of Theorem 2.10) that if an equation involving additive functions on B^m to B is identically satisfied, then the corresponding equation involving their extensions is also identically satisfied. Such a statement is also true (Theorem 2.11) of certain implications between two equations, and it yields several interesting results.

The purpose of the present note is to give a direct proof of an extension of that Jónsson-Tarski's Theorem 2.10. This extension (Theorem II) consists in getting the conclusion under a weaker hypothesis on the functions $g_0, \cdots, g_{m-1}$, namely the monotonicity instead of membership in the set ϕ.

* Received March 31, 1950.

Otherwise our statement (Theorem II) is as Jónsson-Tarski's Theorem 2.10. (It must be pointed out that in that same paper it is also shown: Theorem 2.10 does not hold with the hypothesis that f is monotonic even when $g_0, \cdots, g_{m-1}$ are additive, and on the other hand, it holds whenever f is monotonic and $g_0, \cdots, g_{m-1}$ are "identity functions" (Theorem 2.9)). Throughout our proof we shall make free use of many of the terminology and notation in Jónsson-Tarski's paper, and we shall continue to refer to its definitions and theorems by using the reference numbers therein.[1]

THEOREM I. *Let* $B = \langle B_0, +, 0, \cdot, 1\rangle$ *be a regular subalgebra of a Boolean algebra* $A = \langle A_0, +, 0, \cdot, 1\rangle$ *and let* m *and* n *be positive integers. Then, if* f *is an additive function on* B^m *to* B *and* $g_0, \cdots, g_{m-1}$ *are monotonic functions on* B^n *to* B *we have*

$$(f[g_0, \cdots, g_{m-1}])^+ = f^+[g_0^+, \cdots, g_{m-1}^+].$$

Proof. First we remark that f^+/C^m is on C^m to C and g_j^+/C^n $(j = 0, \cdots, m-1)$ are on C^n to C. The inclusion

1) $f^+[g_0^+, \cdots, g_{m-1}^+] \leq (f[g_0, \cdots, g_{m-1}])^+$ is easily checked: Using the definition of composition, 2.1 and 2.2, the hypothesis on f together with 2.4, and then the remark that $x \geq y^0, \cdots, y^{m-1} \varepsilon C^n$ implies $x \geq y^0 + \cdots + y^{m-1} \varepsilon C^n$ together with the monotonicity of f^+ and g_j^+ $(j = 0, \cdots, m-1)$, we have for every $x \varepsilon A^n$

$$\begin{aligned} f^+[g_0^+, \cdots, g_{m-1}^+](x) &= f^+(g_0^+(x), \cdots, g_{m-1}(x)) \\ &= f^+\Big(\sum_{x \geq y^0 \varepsilon C^n} g_0^+(y^0), \cdots, \sum_{x \geq y^{m-1} \varepsilon C^n} g_{m-1}^+(y^{m-1})\Big) \\ &= \sum_{x \geq y^0 \varepsilon C^n} \cdots \sum_{x \geq y^{m-1} \varepsilon C^n} f^+(g_0^+(y^0), \cdots, g_{m-1}^+(y^{m-1})) \\ &\leq \sum_{x \geq y \varepsilon C^n} f^+(g_0^+(y), \cdots, g_{m-1}^+(y)). \end{aligned}$$

By 2.2 the last sum is $\sum_{x \geq y \varepsilon C^n} \prod_{\langle g_0^+(y), \ldots, g_{m-1}^+(y)\rangle \leq z \varepsilon B^m} f(z)$ and it is included in $\sum_{x \geq y \varepsilon C^n} \prod_{y \leq z \varepsilon B^n} f(g_0(z), \cdots, g_{m-1}(z))$, since every factor of each product of this sum is a factor of the corresponding product of the above sum: f is monotonic, and $\langle g_0^+(y), \cdots, g_{m-1}^+(y)\rangle \leq \langle g_0(z), \cdots, g_{m-1}(z)\rangle \varepsilon B^m$ whenever $y \leq z \varepsilon B^n$ because of the monotonicity of g_j^+ $(j = 0, \cdots, m-1)$. Now

[1] The results of the present note were originated from and reported to Professor Tarski's seminar on Topics in algebra and metamathematics at the University of California, Berkeley.

$$\sum_{x \geq y \varepsilon C^n} \prod_{y \leq z \varepsilon B^n} f(g_0(z), \cdots, g_{m-1}(z))$$
$$= \sum_{x \geq y \varepsilon C^n} \prod_{y \leq z \varepsilon B^n} f[g_0, \cdots, g_{m-1}](z) = (f[g_0, \cdots, g_{m-1}])^+(x),$$

by the definition of composition and then by 2. 1.

From this proof of 1) it follows that, for every $x \varepsilon C^n$ such an inclusion holds even if f is monotonic not additive. In this case the sequence of equalities and inclusions yielding 1) will, essentially, begin after the first inclusion above, and there the additivity of f does not play any role.

Next, we prove for f monotonic (not necessarily additive) the inclusion

$$2) \qquad (f([g_0, \cdots, g_{m-1}])^+ \leq f^+[g_0^+, \cdots, g_{m-}^+].$$

From the definition of composition, 2. 1 and 2. 2 it is clear that it will be sufficient to show that for every $y \varepsilon C^n$ we have

$$(f[g_0, \cdots, g_{m-1}])^+(y) \leq f^+(g_0^+(y), \cdots, g_{m-1}^+(y)).$$

By 2. 2 and, again by the definition of composition, this inclusion will be established, for any $y \varepsilon C^n$, if we prove that to each factor of

$$\prod_{\langle g_0^+(y), \cdots, g_{m-1}^+(y)\rangle \leq z' \varepsilon B^m} f(z')$$

there is at least one factor of $\prod_{y \leq z'' \varepsilon B^n} f(g_0(z''), \cdots, g_{m-1}(z''))$ which is included in it.

Since f is monotonic it is now sufficient to show that to each $z' \varepsilon B^m$ for which $\langle g_0^+(y), \cdots, g_{m-1}^+(y)\rangle \leq z'$, there is some $z'' \varepsilon B^n$ having the properties I) $y \leq z''$, II) $\langle g_0(z''), \cdots, g_{m-1}(z'')\rangle \leq z'$.

To do this let $z' = \langle z'_0, \cdots, z'_{m-1}\rangle \varepsilon B^m$ and let us remark that, by 2. 2. our hypothesis means

$$\prod_{y \leq z \varepsilon B^n} g_j(z) \leq z'_j \varepsilon B \qquad (j = 0, \cdots, m-1).$$

First, we have that *for any* $j = 0, \cdots, m-1$ *there is* $z''^j \varepsilon B^n$ *such that at same time* $y \leqq z''^j$ *and* $g_j(z''^j) \leqq z'_j$. This is true since z'_j being open and including a product of closed elements, it will include (by 1. 21, (iv)) some finite product $\prod_{k=0}^{r_j} g_j(z^k)$ of such closed (and open) factors:

$$\prod_{k=0}^{r_j} g_j(z^k) \leq z'_j \qquad (k = 0, \cdots, r_j)$$

with $y \leq z^k \varepsilon B^n$. Now, putting $z''^j = \prod_{k=0}^{r_j} z^k$, we will have not only $y \leq z''^j \varepsilon B$

but also $g_j(z''^j) \leq z'_j$, since $g_j(z''^j) \leq \prod_{k=0}^{r_j} g_j(z^k)$ because of the monotonicity of g_j. As second and final step it is easily seen that $z'' = \prod_{j=0}^{m-1} z''^j$ is an element of B^n having the properties I) and II) above: $y \leq z''$ since $y \leq z''^j$ for every $j = 0, \cdots, m-1$; and $g_j(z'') \leq z'_j$ for every $j = 0, \cdots, m-1$, since g_j being monotonic we have $g_j(z'') \leq g_j(z''^j) \leq z'_j$.

The proof of 2) is now complete. From 1) and 2) the theorem follows.

Remark. As a consequence of the preceding proof of 2) and of the comment at the end of the proof of 1) we have

$$(f[g_0, \cdots, g_{m-1}])^+(y) = f^+[g_0^+, \cdots, g_{m-1}^+](y)$$

whenever $f, g_0, \cdots, g_{m-1}$ are monotonic and $y \in C^n$.

THEOREM II. *Let $B = \langle B_0, +, 0, \cdot, 1\rangle$ be a regular subalgebra of a Boolean algebra $A = \langle A_0, +, 0, \cdot, 1\rangle$, let m and n be positive integers and let ϕ be the smallest set having the two properties:*

i) *to include all additive functions on B^t to B for any t (integer positive).*

ii) *to be closed in respect to the operation of composition (of functions).*

Then, if $g_0, \cdots, g_{m-1}$ are monotonic functions on B^n to B and $f \in \phi$ is a function on B^m to B, we have

$$(f[g_0, \cdots, g_{m-1}])^+ = f^+[g_0^+, \cdots, g_{m-1}^+].$$

Proof. Remark first that the operation of composition is associative and that for a function f on B^m to B to verify the hypothesis it is necessary (and sufficient) that non negative integers k and $r_0 = 0, r_1, \cdots, r_k$ exist such that

$$f = h_0^0[h_0^1, \cdots, h_{r_1}^1] \cdots [h_0^k, \cdots, h_{r_k}^k]$$

for some additive functions h_j^k $(j = 0, \cdots, r_k)$ on B^m to B and h_j^i $(i = 0, \cdots, k-1; j = 0, \cdots, r_i)$ on $B^{r_{i+1}}$ to B.

Put $h = h_0^0$. If we have $k = 0$ in the equality above, then f is just the additive function h, and we have the desired conclusion from Theorem I. We prove by induction for $k \neq 0$. Putting

$$f'_j = h_j^1[h_0^2, \cdots, h_{r_2}^2] \cdots [h_0^k, \cdots, h_{r_k}^k] \qquad (j = 0, \cdots, r_1)$$

we have

$$(f'_j[g_0, \cdots, g_{m-1}])^+ = f_j^+[g_0^+ \cdot \cdots, g_{m-1}] \qquad (j = 0, \cdots, r_1)$$

as induction hypothesis. On the other hand, since f'_j and also $f'_j[g_0, \cdots, g_{m-1}]$ $(j = 0, \cdots, r_1)$ are monotonic and h is additive, we have

$$h^+[f'^+_0, \cdots, f'^+_{r_1}] = (h[f'_0, \cdots, f'_{r_1}])^+$$

and also

$$\begin{aligned}(h[f'_0[g_0, \cdots, g_{m-1}], \cdots, f'_{r_1}[g_0, \cdots, g_{m-1}]])^+ \\ = h^+[(f'_0[g_0, \cdots, g_{m-1}])^+, \cdots, (f'_{r_1}[g_0, \cdots, g_{m-1}])^+],\end{aligned}$$

by Theorem I. Hence,

$$\begin{aligned}(f[g_0, \cdots, g_{m-1}])^+ &= ((h[f'_0, \cdots, f'_{r_1}])[g_0, \cdots, g_{m-1}])^+ \\ &= (h[f'_0[g_0, \cdots, g_{m-1}], \cdots, f'_{r_1}[g_0, \cdots, g_{m-1}]]^+ \\ &= h^+[(f'_0[g_0, \cdots, g_{m-1}])^+, \cdots, (f'_{r_1}[g_0, \cdots, g_{m-1}])^+] \\ &= h^+[f'^+_0[g_0^+, \cdots, g_{m-1}^+], \cdots, f'^+_{r_1}[g_0^+, \cdots, g_{m-1}^+]] \\ &= (h^+[f'^+_0, \cdots, f'^+_{r_1}])[g_0^+, \cdots, g_{m-1}^+] \\ &= (h[f'_0, \cdots, f'_{r_1}])^+[g_0^+, \cdots, g_{m-1}^+] = f^+[g_0^+, \cdots, g_{m-1}^+].\end{aligned}$$

Thus Theorem II is proved.

UNIVERSITY OF CALIFORNIA,
BERKELEY, CALIFORNIA.

CRITÈRES DE COMPACITÉ DANS LES ESPACES FONCTIONNELS GENERAUX.*

By A. GROTHENDIECK.

1. Introduction.[1] Outre la notion usuelle de compacité (cf. N. Bourbaki [1]), qui est apparue come étant la seule qui soit vraiment fondamentale, on rencontre néanmoins au moins deux autres notions étroitement apparentées, mais de "caractère dénombrable," et qui se sont révélées indispensables dans plusieurs questions qui à priori n'impliquent aucune considération de dénombrabilité.

Nous dirons qu'une partie A d'un espace topologique est ***relativement semi-compacte*** (resp. *semi-compacte*) si toute suite extraite de A admet une valeur d'adhérence (resp. qui appartienne à A). A sera dite ***strictement relativement semi-compacte*** (resp. ***strictement semi-compacte***) si de toute suite extraite de A on peut extraire une suite converge (resp. qui converge vers un élément de A).

Dans le présent travail, nous étudions des cas étendus où la semi-compacité relative entraine déjà la compacité relative ou la stricte semi-compacité relative. Nous nous y plaçons surtout dans des espaces du type $C_{\mathfrak{S}}(E,F)$, espace des applications continues d'un espace topologique E dans un espace uniforme séparé F, muni de la topologie de la convergence uniforme sur un ensemble $\mathfrak{S}$ de parties de E recouvrant E (pour ces notions fondamentales d'Analyse Fonctionnelle, cf. N. Bourbaki [3]). Les résultats obtenus valent manifestement pour les sous-espaces fermés de tels espaces, ce qui permet de les appliquer à des espaces vectoriels localement convexes généraux.

Dans 2, nous donnons quelques généralités sur les diverses notions de compacité envisagées, destinées surtout à prémunir le lecteur contre certaines

* Received December 27, 1950.

[1] Outre le théorème 5 et la proposition 7 du présent travail, citons notamment encore deux propositions s'appuyant de façon essentielle sur le théorème d'Eberlein (cf. proposition 2 ci-dessous): 1) l'enveloppe convexe fermée d'une partie faiblement relativement compacte d'un espace de Banach (par exemple) est faiblement compacte; 2) Le produit de deux fonctions faiblement presque-périodiques sur un semi-groupe est faiblement presque-périodique (cf. [6]); plus généralement, si E est une algèbre normée complète s'identifiant à l'espace des fonctions complexes continues sur un espace compact, le produit de deux parties faiblement compactes A et B de E (ensemble des xy avec $x \varepsilon A$ et $y \varepsilon B$) est faiblement relativement compact.

erreurs assez naturelles, plutôt que de repéter les développements bien connus et triviaux sur le sujet (tels que: l'image continue d'un espace semi-compact est semi-compact, etc.). Dans 3. nous étudions des cas où la semi-compacité relative entraine la compacité relative, le resultat le plus important est le théorème 2 (et le théorème 1 qui est un corollaire) ; nous y donnons en même temps un critère de compacité relative qui semble à priori encore bien plus faible que la semi-compacité relative. En outre, M. J. Dieudonné a bien voulu me communiquer un autre cas non trivial et très simple où la semi-compacité relative entraine la compacité relative (théorème 3), critère qui ne sera pas essentiel par la suite mais a son intérêt propre dans l'ordre d'idées de ce travail. Dans 4, nous appliquons les résultats obtenus aux espaces localement convexes, en donnant notamment à un classique théorème d'Eberlein pour les espaces de Banach (généralisé par J. Dieudonné et L. Schwartz [5] aux espaces ($\mathfrak{F}$)) toute la généralité qui lui appartient. (Ce théorème a été d'ailleurs le point de départ du présent travail). Dans 5. nous établissons un cas non classique ou la semi-compacité relative entraine la *stricte* semi-compacité relative (th. 4) ; ce résultat est d'ailleurs essentiel pour la suite (th. 5) ; la proposition 5 se réduit à une systématisation de réflexions classiques. Nous appliquons ensuite les resultats précédents à la détermination des parties faiblement relativement compactes de l'espace de Banach $C^\infty(E)$ de toutes les fonctions continues et bornées sur un espace topologique E; le critère obtenu donne par exemple immédiatement le résultat suivant. Une fonction faiblement presque périodique à gauche sur un semi-groupe est aussi faiblement presque périodique à droite. Enfin dans 7. nous généralisons le théorème 6 pour obtenir dans les espaces localement convexes un critère de relative compacité faible, approfondissant de beaucoup les résultats de 4, et qui ne semble pas connu même pour les espaces de Banach.

2. Généralités. Il n'est peut-être pas inutile de rappeler quelles implications on peut ou ne peut pas affirmer entre les diverses notions de compacité envisagées, et quelles simplifications se produisent dans quelques cas classiques. Il est évident que la compacité et la semi-compacité stricte entraînent chacune la semi-compacité, de même pour les notions "relatives" correspondantes. Mais on n'a dans le cas général aucune autre implication entre ces trois couples de notions, car il est bien connu qu'un espace strictement semi-compact peut être non compact (exemple: espace des nombres ordinaux de seconde classe, avec la topologie usuelle) et un espace compact peut ne pas être strictement semi-compact (exemple: produit topologique d'une famille non dénombrable d'intervalles compacts).—D'autre part, il est évident

que chacune des trois notions de compacité entraine la notion "relative" correspondante, et que la réciproque est tout à fait fausse. On fera attention ici que pour qu'une partie A d'un espace E soit relativement compacte, il faut et il suffit par définition que son adhérence soit compacte, mais qu'il n'est est plus de même pour la semi-compacité relative et la stricte semi-compacité relative. La condition est évidemment encore suffisante, mais on peut trouver une partie A strictement semi-compacte d'un espace séparé E, dont l'adhérence ne soit pas même semi-compacte. En d'autres termes, il existe un espace séparé non semi-compact E dans lequel une partie strictement semi-compacte A soit dense. Soit en effet X un espace séparé qui soit strictement semi-compact et localement compact mais non compact (par exemple l'espace des nombres ordinaux de seconde classe), a son "point à l'infini," $Y = X \cup (a)$. Pour tout entier naturel n, soit Y_n un exemplaire homéomorphe de Y (X_n correspondant à X et a_n à a) ; supposons les Y_n disjoints et soit b un élément qui n'appartienne à aucun des Y_n. Sur l'ensemble $E = (b) \cup \bigcup_n Y_n$, considérons la topologie dont les ouverts sont les parties qui coupent chaque Y_n suivant un ouvert, et qui, s'ils contiennent b, contiennent aussi les X_n à partir d'un rang assez élevé. On vérifie trivialement les axiomes des ouverts (Bourbaki (1)), et que E est séparé; E n'est pas semi-compact, car il est manifeste que la suite $(a_n)_n$ n'a pas de point adhérent. D'autre part, $A = (b) \cup \bigcup_n X_n$ est partout dense et strictement semi-compact, comme on vérifie aussitôt.

Rappelons enfin que dans un espace métrique, les trois notions de compacité sont équivalentes, ainsi que les notions "relatives" correspondantes. Un autre résultat intéressant, qui nous sera utile par la suite, est le théorème d'A. Weil [10] ; une partie relativement semi-compacte d'un espace uniforme séparé est précompacte. En particulier, dans un espace uniforme séparé et complet, compacité (relative) et semi-compacité (relative) sont la même chose.

La topologie faible d'un espace de Banach ou d'un espace localement convexe quelconque est un exemple d'un topologie en général ni complète ni métrisable, et où les critères de relative compacité qu'on vient de rappeler ne s'appliquent pas tels quels. Plus généralement, il en est ainsi dans les espaces d'applications continues d'un espace topologique dans un autre, muni de la topologie de la convergence simple par exemple. Pourtant un théorème d'Eberlein pour les espaces de Banach, et les résultats de G. Köthe sur ses "espaces parfaits" (cf. G. Köthe [8]) montrent que dans ces espaces, munis de la topologie faible, on a encore identité entre parties relativement semi-compactes et relativement compactes. D'autre part, dans les espaces de Banach encore, un théorème de Šmulian (généralisé aux espaces ($\mathfrak{F}$) dans [5]) affirme

l'identité pour la topologie faible entre parties strictement relativement semi-compactes et parties relativement compactes. Ce sont ces résultats qui nous ont guidé et que nous allons généraliser et préciser.

3. Semi-compacité et compacité. Soit E un espace topologique, F un espace uniforme séparé; nous désignons par $C(E,F)$ l'espace des applications continues de E dans F, par $\mathfrak{F}(E,F)$ l'espace de toutes les applications de E dans F, et, si $\mathfrak{S}$ est un ensemble de parties de E, par $\mathfrak{T}_{\mathfrak{S}}$ la structure uniforme sur $C(E,F)$ et $\mathfrak{F}(E,F)$ de la convergence uniforme sur les éléments de $\mathfrak{S}$ (cf. [3]) ; munis de cette structure les espaces précédents seronts désignés par $C_{\mathfrak{S}}(E,F)$ resp. $\mathfrak{F}_{\mathfrak{S}}(E,F)$. Si F est séparé et si $\mathfrak{S}$ recouvre E (ce que nous supposons par la suite) ces espaces sont séparés, et si de plus F est complet, il en est même de $\mathfrak{F}_{\mathfrak{S}}(E,F)$, mais en général $C_{\mathfrak{S}}(E,F)$ n'est pas complet. On vérifie aussitôt que si un filtre de Cauchy dans $\mathfrak{F}_{\mathfrak{S}}(E,F)$ converge pour la topologie de la convergence simple, il converge pour $\mathfrak{T}_{\mathfrak{S}}$, d'où suit que pour qu'une partie A de $C_{\mathfrak{S}}(E,F)$ ait une adhérence complète dans cet espace, il faut et il suffit que tout filtre de Cauchy sur A converge en chaque point vers une application *continue* de E dans F. Alors l'adhérence $\bar{A}$ de A pour $\mathfrak{T}_{\mathfrak{S}}$ sera à fortiori complète pour toute $\mathfrak{T}_{\mathfrak{S}'}$, avec $\mathfrak{S}' \supset \mathfrak{S}$. En particulier si A a une adhérence complète dans l'espace $C_s(E,F)$ muni de la structure $\mathfrak{T}_s$ de la convergence simple, il en sera de même à fortiori pour toute $\mathfrak{T}_{\mathfrak{S}}$. A fortiori, si A est relativement compacte dans $C_s(E,F)$, l'adhérence de A dans $C_{\mathfrak{S}}(E,F)$ est complète quel que soit l'ensemble de parties $\mathfrak{S}$.

Enfin, remarquons encore que le théorème de Tychonoff donne immédiatement: Pour qu'une partie de $C_s(E,F)$ soit relativement compacte, il faut et il suffit que 1°) elle le soit dans le produit topologique $\mathfrak{F}_s(E,F)$, c'est-à-dire que pour tout $x \varepsilon E$, l'ensemble des $f(x)$, où $f \varepsilon A$, soit relativement compact dans F; et 2°) que l'adhérence de A dans $C_s(E,F)$ soit la même que dans $\mathfrak{F}_s(E,F)$, c'est-à-dire que toute application de E dans F qui est limite simple d'applications éléments de A, soit continue.

Ces remarques interviennent dans diverses questions d'Analyse Fonctionnelle, et seront essentielles pour la compréhension de la suite.

THÉORÈME 1. *Soit E un espace semi-compact, F un espace uniforme séparé, $\mathfrak{S}$ un ensemble de parties de E recouvrant E. Si dans F toute partie relativement semi-compacte est relativement compacte (en particulier, si F est complet), alors il en est de même dans l'espace $C_{\mathfrak{S}}(E,F)$.*

Toute partie relativement semi-compacte A de $C_{\mathfrak{S}}(E,F)$ est précompacte (cf. ci-dessus 2.) ; il suffit de montrer que son adhérence est complète, et a

fortiori, d'après nos remarques précédentes, que A est relativement compact pour la topologie de la convergence simple. Comme A est évidemment relativement semi-compact pour cette dernière topologie (puisque $\mathfrak{S}$ recouvre E) on est ramené au cas de la topologie $\mathfrak{T}_s$. Mais ce cas est inclus dans le théorème-clef suivant:

THÉORÈME 2. *Soit E un espace semi-compact, F un espace complètement régulier, A une partie de $C_s(E,F)$, E_1 une partie dense de E.*

1°) *Si dans F toute partie relativement semi-compacte est relativement compacte, alors les conditions suivantes sur A sont toutes équivalentes:*

a) *A est relativement compact;*

b) *A est relativement semi-compact;*

c) *pour toute suite (f_n) extraite de A et toute suite (x_i) extraite de E_1, il existe une application continue f de l'adhérence K de l'ensemble des x_i dans F, telle que pour tout $x \in K$, $f(x)$ soit adhérent à la suite des $f_n(x)$. Et pour tout $x \in \complement E_1$, l'ensemble des $f(x)$ avec $f \in A$ est relativement semi-compact;*

d) *pour toute suite (f_n) extraite de A et toute suite (x_i) extraite de E_1, il existe un $X \in F$ qui soit point doublement adhérent à la suite double $(f_n(x_i))$ (par quoi nous entendons que tout voisinage de X rencontre une infinité de lignes et une infinité de colonnes de la suite double chacune en une infinité de termes). Et pour tout $x \in \complement E_1$, l'ensemble des $f(x)$ avec $f \in A$ est relativement semi-compact.*

2°) *De toutes façons (sans plus faire sur F la restriction de la première partie de l'énoncé) chacune des conditions qui précèdent est suffisante pour assurer que toute application de E dans F qui est limite simple d'applications éléments de A est continue; les deuxièmes parties des conditions* c) *et* d) *peuvent être omises.*

Enfin, moyennant la première partie de la condition d), *même si on ne suppose plus que E est semi-compact, toute application de E dans F qui est limite simple d'applications éléments de A est continue.*

Démonstration. On a de toutes façons manifestement a)$\rightarrow$b)$\rightarrow$c); montrons que si E est semi-compact, c) entraine d); il suffit de montrer que la première partie de la condition c) entraine la première partie de la condition d). Soit en effet, avec les notations de d), $x_0 \in E$ adhérent à la suite (x_i) et soit f l'application stipulée dans c), relative aux suites (x_i) et (f_n); je dis que $f(x_0)$ est doublement adhérent à la suite double $(f_n(x_i))$. En effet,

s'il existait un voisinage ouvert V de $f(x_0)$ tel que sauf pour un nombre fini d'indices i l'on ait "$f_n(x_i) \varepsilon \complement V$ pour $n \geqq n_0(i)$," on aurait,, sauf pour un nombre fini d'indices: $f(x_i) \varepsilon \complement V$, d'où $f(x_0) \varepsilon \complement V$ ce qui est absurde; et s'il existait un voisinage ouvert V de $f(x_0)$ tel que sauf pour un nombre fini d'indices n l'on ait "$f_n(x_i) \varepsilon \complement V$ pour $i \geqq i_0(n)$," on aurait sauf pour un nombre fini d'indices $f_n(x_0) \varepsilon \complement V$, d'où $f(x_0) \varepsilon \complement V$, ce qui est encore absurde.— Pour prouver la première partie du théorème, tout revient donc à prouver que d) entraine a). Mais de d) resulte manifestement que pour *tout* $x \varepsilon E$ l'ensemble des $f(x)$, avec $f \varepsilon A$ est une partie relativement semi-compacte de F, donc relativement compacte en vertu de l'hypothèse sur F. En tenant compte d'une remarque faite plus haut, tout revient donc à montrer que toute application de E dans F qui est limite simple d'applications éléments de A est continue. Cela est inclus dans la deuxième partie du théorème, cette deuxième partie revenant manifestament à prouver que si E est un espace topologique quelconque, et F complètement régulier, alors la première condition énoncée dans d) est suffisante pour assurer que toute application f de E dans F qui est limite simple d'applications éléments de A est continue (E_1 désignant une partie dense dans E).

F étant complèment régulier, sa topologie peut être considérée comme la moins fine de celles qui rendent continues certaines fonctions numériques ϕ_i sur F ([2], page 11, proposition 4). On voit alors qu'on peut se ramener au cas où F est la droite numérique, la continuité de f équivalant en effet à la continuité de chacun des fonctions numériques $\phi_i \circ f$ sur E (d'autre part $\phi_i \circ f$ est limite simple de fonctions $\phi_i \circ g$ où g parcourt A, et l'ensemble de ces $\phi_i \circ g$ jouit manifestement des propriétés envisagées pour A lui-même). Supposons donc que F soit la droite numérique; on sait que pour démontrer la continuité de f, il suffit de montrer que l'on a $\lim\limits_{\substack{x \to x_0 \\ x \varepsilon E_1}} f(x) = f(x_0)$ pour tout $x_0 \varepsilon E$ ([1] page 38, th. 1). Nous démontrerons cette relation par l'absurde, en reprenant une idée d'Eberlein. Supposons donc qu'il existe un $x_0 \varepsilon E$ et un $\alpha > 0$ tels que pour tout voisinage V de x_0, il existe un $x \varepsilon V \cap E_1$ tel que $|f(x) - f(x_0)| \geqq \alpha$. On pourrait alors par récurrence construire deux suites d'éléments de A et de E_1 respectivement, (f_i) et (x_i), telles que l'on ait (on suppose les suites construites déjà jusqu'aux termes de rang $n-1$):

a) $|f_n(x_i) - f(x_i)| \leqq 1/n$ $(0 \leqq i \leqq n-1)$ (cela est possible, f étant limite simple d'éléments de A).

b) $|f_i(x_n) - f_i(x_0)| \leqq 1/n$ $(0 \leqq i \leqq n)$.

c) $|f(x_n) - f(x_0)| \geqq \alpha$ (ce que est encore possible par hypothèse).

Il existe un point z doublement adhérent à la suite double $(f_i(x_j))$, et comme pour i constant la suite $(f_n(x_i))$ tend vers $f(x_i)$ en vertu de a) et la suite $(f_i(x_n))$ vers $f_i(x_0)$ en vertu de b), z est adhérent à chancune des suites $(f(x_n))$ et $(f_n(x_0))$. Or, en vertu de la première des inégalités a), la deuxième suite tend vers $f(x_0)$; on a donc $z = f(x_0)$, et $f(x_0)$ serait valeur d'adhérence de la première suite, contrairement aux inégalités c). CQFD.

COROLLAIRE 1. *L'énoncé du théorème* 1 *reste valable si on suppose E localement compact, ou métrique, et plus généralement si toute application de E dans F dont les restrictions aux parties semi-compactes de E sont continues, est continue.*

En effet, on se ramène évidemment à montrer que si A est relativement semi-compact pour la topologie de la convergence simple, toute limite simple d'applications éléments de A est continue, ce qui est déjà une conséquence du théorème 1.

Remarque 1. La démonstration du théorème 2 met en évidence que si on suppose la topologie de F définie comme la moins fine des topologies rendant continues certaines applications ϕ_i de F dans des espaces complètement réguliers F_i, alors les critères énoncés dans le théorème 2 équivalent à ceux qu'on en déduit en supposant que les hypothèses envisagées sont vérifiées, non pour A lui-même à priori, mais pour chacun des ensembles $A_i \subset C(E, F_i)$ (où pour tout i, on désigne par A_i l'ensemble des $\phi_i \circ f$ où f parcourt A).

Remarque 2. Supposons que la suite double (x_{ij}) prenne ses valeurs dans un espace métrique F, et y soit relativement compacte. Alors on vérifie que la non-existence d'un point doublement adhérent à la suite double implique l'existence d'une "suite double extraite" $(x_{i_\alpha j_\beta}) = (y_{\alpha,\beta})$, telle que $\lim_\alpha . \lim_\beta . y_{\alpha\beta}$ et $\lim_\beta . \lim_\alpha . y_{\alpha\beta}$ existent tous deux et soient distincts. En effet, l'application du procédé diagonal permet de construire une suite d'indices (i_α) telle que $\lim_{\alpha\to\infty} . x_{i_\alpha, j}$ existe pour tout j. Une seconde application du procédé diagonal permet d'obtenir une suite d'indices j_β telle que $\lim_{\beta\to\infty} . x_{i_\alpha, j_\beta}$ existe pour tout i_α, et que $\lim_\beta . (\lim_\alpha . x_{i_\alpha, j_\beta})$ existe. Enfin on peut supposer en extrayant encore au besoin une suite partielle de la suite (i_α), que $\lim_\alpha . (\lim_\beta . x_{i_\alpha, j_\beta})$ existe. Mais les deux limites doubles $\lim_\alpha . (\lim_\beta . x_{i_\alpha, j_\beta})$ et $\lim_\beta . (\lim_\alpha . x_{i_\alpha, j_\beta})$ ne peuvent être égales, car leur valeur commune serait manifestement un point doublement adhérent à la suite double (x_{i_α, j_β}).

D'autre part, une suite double telle que $\lim_i . (\lim_j . x_{i,j})$ et $\lim_j . (\lim_i . x_{ij})$

existent et soient distincts n'a manifestement pas de point doublement adhérent (car un tel point devrait être identique à chacune de ces limites doubles). Il suit aussitôt le

COROLLAIRE 2. *Soit E un espace semi-compact, F un espace métrique, A un ensemble d'applications continues de E dans F tel que l'ensemble des $f(x)$ où $f \varepsilon A$ et $x \varepsilon E$, soit relativement compact. Pour que A soit relativement compact dans $C_s(E, F)$, il faut et il suffit qu'il n'existe pas de suite (x_i) extraite de E et de suite (f_j) extraite de A, telles que $\lim_i \lim_j f_j(x_i)$ et $\lim_j \lim_i f_j(x_i)$ existent tous deux et soient distincts. La condition subsiste si on assujettit la suite (x_{ij}) à être extraite d'une partie partout dense fixe E_1 de E. Et cette condition reste suffisante pour assurer que A est relativement compact, même si E n'est plus supposé semi-compact.*

De la démonstration du théorème 2, ou du théorème 2 directement, on déduit immédiatement le résultat suivant:

PROPOSITION 1. *Soit E un espace semi-compact, E un espace complètement régulier, A une partie relativement compacte de l'espace $C_s(E, F)$. Alors A est encore relativement compact dans l'espace $C_s(\bar{E}, F)$, où $\bar{E}$ est l'espace obtenu en munissant E de la topologie la moins fine rendant continues les applications éléments de A. En particulier, toute application f de E dans F que est limite simple d'applications éléments de A, est encore continue au sens de la topologie de $\bar{E}$* (c'est à dire que, pour tout $x_0 \varepsilon E$ et tout voisinage V de $f(x_0)$ dans F, il existe un nombre fini d'éléments $f_i \varepsilon A$ et des ouverts Ω_i dans F, tels que $f_i(x_0) \varepsilon \Omega_i$ pour tout i, et que $f_i(x) \varepsilon \Omega_i$ pour tout i entraine $f(x) \varepsilon V$.[2]

Signalons pour être complet un autre cas intéressant et non classique où la semi-compacité relative entraine la compacité relative, qui m'a été signalé par M. J. Dieudonné:

THÉORÈME 3. *Soit E un espace complètement régulier dont la topologie $\mathfrak{T}$ soit plus fine qu'une certaine topologie métrisable $\mathfrak{T}_0$. Alors dans E les parties (relativement) compactes, (relativement) semi-compactes et (relativement) strictement semi-compactes sont identiques, et leur topologie métrisable.*

[2] En fait, cet énoncé est loin d'être profond, du moins si E est compact. On peut en effet montrer alors par voie directe le résultat bien moins restrictif: Si E est *compact*, F un espace topologique *séparé quelconque*, A un ensemble *quelconque* d'applications continues de E dans F, alors toute application *continue* de E dans F qui est limite simple d'applications éléments de A, est déjà continue quand on munit E de la topologie la moins fine rendant continues les applications éléments de A.

Il suffit de montrer que toute partie relativement semi-compacte A est relativement compacte. Car alors, son adhérence $\bar{A}$ étant compacte et la topologie induite par $\mathfrak{T}_0$ sur $\bar{A}$ étant séparée et moins fine que celle induite par $\mathfrak{T}$, elle doit lui être identique, d'où suit que $\bar{A}$ est métrisable et *strictement* semi-compact. Tout revient donc à montrer que tout ultra-filtre ϕ sur A converge vers quelque $x_0 \in E$. Mais A étant aussi relativement semi-compact pour $\mathfrak{T}_0$ qui est métrisable, A est relativement compact pour $\mathfrak{T}_0$, donc ϕ converge pour $\mathfrak{T}_0$ vers un $x_0 \in E$. Tout revient à montrer que la convergence a lieu aussi au sens de $\mathfrak{T}$, donc (cf. [2] p. 11, proposition 9) que pour toute fonction numérique continue f sur E, $f(x)$ converge vers $f(x_0)$ suivant le filtre ϕ. Soit $\mathfrak{T}_f$ la topologie la moins fine sur E rendant continues f et l'application identique de E sur E muni de $\mathfrak{T}_0$, cette topologie est métrisable, plus fine que $\mathfrak{T}_0$ et moins fine que $\mathfrak{T}$. A est donc aussi relativement semi-compact pour $\mathfrak{T}_f$, donc relativement compact pour cette topologie, ϕ tend donc vers une limite $y \in E$ au sens de $\mathfrak{T}_f$, et on a forcément $y = x_0$ puisque $\mathfrak{T}_f$ est plus fine que $\mathfrak{T}_0$. Il suit bien que $f(x)$ tend vers $f(x_0)$ suivant le filtre ϕ, CQFD.

COROLLAIRE. *Soit E un espace topologique, F un espace métrique, $\mathfrak{S}$ un ensemble de parties de E recouvrant E. Supposons qu'il existe une suite d'ensembles éléments de $\mathfrak{S}$ dont la réunion soit partout dense dans E (en particulier, il suffit qu'il existe dans E une suite partout dense). Alors dans $C_{\mathfrak{S}}(E,F)$ les parties (relativement) compactes, (relativement) semi-compactes et (relativement) strictement semi-compactes sont identiques.*

Remarquons que le théorème 3 aurait pu se démontrer aussi rapidement sans l'aide des ultra-filtres, en montrant directement que sur l'adhérence des parties relativement semi-compactes, les topologies $\mathfrak{T}$ et $\mathfrak{T}_0$ sont identiques. Mais la méthode employée montre plus généralement que si on considère un ensemble de topologies $\mathfrak{T}_i$ sur E où les parties (relativement) semi-compactes soient (relativement) compactes, et si cette famille de topologies admet un plus petit élément $\mathfrak{T}_0$ séparé, alors la borne supériere $\mathfrak{T} = \operatorname{Sup}_i \mathfrak{T}_i$ satisfait à la même hypothèse que les $\mathfrak{T}_i$.

4. Appplications aux espaces vectoriels localement convexes. Les théorèmes 1 et 2 s'appliquent aux sous-espaces fermés d'espaces $C_{\mathfrak{S}}(E,F)$. De manière générale, l'application du théorème 1 peut se présenter ainsi: On donne un ensemble B d'applications d'un ensemble E dans un espace uniforme séparé F dont les parties relativement semi-compactes soient relativement compactes, examiner s'il en est de même dans B muni d'une topologie

$\mathfrak{T}_{\mathfrak{S}}$. On pourra l'affirmer dès qu'on aura trouvé sur E, pour toute partie relativement semi-compacte A de B, une topologie rendant continues les applications éléments de A, et assez peu fine pour que toute application de E dans F dont les restrictions aux parties semi-compactes de E sont continues, et qui soit par ailleurs limite au sens de $\mathfrak{T}_{\mathfrak{S}}$ d'applications éléments de A, soit élément de B. Remarque analogue pour l'application du théorème 2, quand $\mathfrak{T}_{\mathfrak{S}}$ est la topologie de la convergence simple, mais alors on a intérêt à prendre sur E une topologie aussi fine que possible donnant encore suffisamment de parties semi-compactes pour que toute application de E dans F dont les restrictions à ces parties sont continues (et de plus limite simple d'éléments de A) soit continue. Ces deux considérations se reflètent exactement dans les deux propositions qui vont suivre.

Si E est un espace vectoriel localement convexe séparé, il peut être considéré comme l'espace des formes linéaires continues sur son dual faible E', muni d'une topologie $\mathfrak{T}_{\mathfrak{S}}$ (théorème de Mackey, cf. [9] et [5]) $\mathfrak{S}$ étant un ensemble de parties convexes et *faiblement compactes* recouvrant E'. D'autre part on peut montrer ([7]) que si E est complet, tout forme linéaire sur E' dont les restrictions aux éléments de $\mathfrak{S}$ sont continues, est faiblement continue, c'est à dire élément de E. Comme par ailleurs toute limite simple d'applications linéaires est linéaire, on obtient en premier lieu la généralisation du théorème d'Eberlein annoncée au début:

Proposition 2. *Si E est un espace localement convexe séparé complet ou seulement complet pour la topologie $\tau(E, E')$ de Mackey associée,* (cf. [5] et [9]) *ses parties relativement semi-compactes et relativement compactes sont identiques* (et ceci d'ailleurs manifestement pour toute topologie localement convexe sur E donnant le même dual).

En second lieu, on a le résultat

Proposition 3. a) *Sous les conditions de la proposition précédente, pour qu'une partie A de E soit faiblement relativement compacte, il faut et il suffit qu'elle soit bornée, et qu'il n'existe pas de suite (x_i) extraite de A et de suite (x'_j) extraite d'une partie faiblement compacte convexe de E', telles que* $\lim_i \lim_j \langle x_i, x'_j \rangle$ *et* $\lim_j \lim_i \langle x_i, x'_j \rangle$ *existent et soient distincts.*

b) *Plus généralement, soit (K_α) une famille de parties convexes de E', relativement faiblement compactes (et non forcément fermées), telle que la famille des adhérences faibles $\overline{K_\alpha}$ engendre algébriquement E', et que E soit complet pour la topologie de la convergence uniforme sur les $\overline{K_\alpha}$. Alors le*

critère précédent de relative faible compacité de A subsiste, si on assujettit la suite (x'_i) à être extraite de quelque K_α.

Remarque 3. On voit facilement que les propositions précédentes valent encore si on suppose non pas E complet, mais seulement ses parties bornées et fermées complètes (il suffit de passer au complété de E pour la topologie donnée ou la topologie $\tau(E, E')$). De façon plus générale encore, les propositions 2 et 3 valent pour une partie particulière A de E, dès qu'on sait que l'enveloppe convexe fermée de A est complète (ne fût-ce que pour $\tau(E, E')$ lorsqu'il s'agit de proposition 2 ou proposition 3a)).—Comme toute partie faiblement compacte de E est forcément complète pour les topologies envisagées, il ne semble pas raisonnable d'espérer généraliser encore ces derniers résultats (mais nous approfondirons encore considérablement la proposition 2 par le théorème 7 plus bas).

Il est d'ailleurs facile de construire un espace vectoriel non complet, hyperplan fortement fermé d'un dual faible d'un espace de Banach par exemple, dans lequel il y ait des parties semi-compactes non relativement compactes. Soit en effet Ω un espace localement compact et semi-compact, mais non compact (par exemple l'espace des nombres ordinaux de seconde classe), soit $\hat{\Omega}$ l'espace compact obtenu par adjonction du "point à l'infini" ϖ. Soit E l'espace des fonctions complexes continues sur $\hat{\Omega}$, muni de la norme de la convergence uniforme, E' son dual (espace des mesures de Radon sur $\hat{\Omega}$). Si on identifie tout point de $\hat{\Omega}$ avec la mass $+1$ placée en ce point, la topologie de $\hat{\Omega}$ s'identifie à la topologie induite par la topologie faible de E'. Il est manifeste que ϖ n'appartient pas au sous-espace fortement fermé engendré par Ω (sa distance à ce dernier est égale à 1), il existe donc un hyperplan fortement fermé V de E' contenant Ω et non ϖ. Dans cet espace (muni de la topologie faible), Ω est semi-compact et non faiblement relativement compact.

La proposition 3 donne un critère pour qu'une suite de E converge faiblement; il faut et il suffit en effet qu'elle soit faiblement relativement compacte, et qu'elle converge sur une partie totale E'_1 de E' (car sur une partie faiblement compacte de E, la topologie $\sigma(E, E')$ coïncide forcément avec la topologie séparée moins fine $\sigma(E, E'_1)$). Nous ne donnons pas l'énoncé explicite, qui de toutes façons pourra beaucoup s'améliorer plus bas. Mais donnons une application immédiate de la proposition 3b), due à ce que le bidual E'' d'un espace E (cf. [5]) est engendré par les adhérences faibles des parties bornées de E:

PROPOSITION 4. *Soit E un espace localement convexe, E' son dual fort*

(cf. [5]) *supposé complet, E'' le dual de E' fort. Pour qu'une partie A de E' soit relativement compacte pour $\sigma(E', E'')$ il faut et il suffit qu'elle soit fortement bornée, et qu'il n'existe pas de suite bornée (x_i) extraite de E et de suite (x'_j) extraite de A, telles que* $\lim_i \lim_j \langle x_i, x'_j \rangle$ *et* $\lim_j \lim_i \langle x_i, x'_j \rangle$ *existent et soient distincts.*

Ici encore, il suffit de supposer seulement que les parties fermées et bornées de E' fort sont complètes. Et on a encore un critère correspondant pour qu'une suite dans E' converge pour $\sigma(E', E'')$: il faut et il suffit qu'elle soit relativement compacte pour cette topologie, et qu'elle converge sur une partie totale de E.—Noter que si on suppose A faiblement relativement compact, il est inutile de supposer E' fort complet (car l'adhérence forte de A sera déjà complète).

5. Critères de semi-compacité stricte. Soit de nouveau E un espace topologique, F un espace uniforme séparé, $\mathfrak{S}$ un ensemble de parties de E recouvrant E. Pour qu'un filtre sur une partie *relativement compacte* A de $C_{\mathfrak{S}}(E, F)$ converge, il faut et il suffit qu'il converge en chaque point d'une partie partout dense E_1 de E (puisque la topologie de la convergence simple sur E_1 est encore séparée sur $C(E, F)$, et moins fine que la topologie $\mathfrak{T}_{\mathfrak{S}}$). Si on suppose seulement A relativement semi-compact, la conclusion subsiste à condition de se borner aux filtres définis par des *suites* (f_n). En effet, cette suite ne peut avoir dans $C_{\mathfrak{S}}(E, F)$ qu'une seule valeur d'adhérence, (définie par ses valeurs sur E_1), et d'autre part on vérifie immédiatement que dans une partie relativement semi-compacte d'un espace topologique séparé C, les suites convergentes sont précisément celles qui ont un seul point adhérent.

PROPOSITION 5. *Soit E un espace topologique, F un espace uniforme séparé, $\mathfrak{S}$ un ensemble de parties de E recouvrant E. Supposons qu'il existe une suite (E_i) de parties de E, dont la réunion soit partout dense, et telle que dans chacun des espaces $C_{\mathfrak{S}_i}(E_i, F)$ ($\mathfrak{S}_i$ désignant la trace de $\mathfrak{S}$ sur E_i), la (semi-)compacité relative d'une partie de l'espace entraine sa stricte semi-compacité relative. Alors il en est de même dans $C_{\mathfrak{S}}(E, F)$.*

Soit en effet (f_n) une suite relativement (semi-)compacte dans $C_{\mathfrak{S}}(E, F)$. Pour tout i, la suite des restrictions des f_n à E_i forme alors une suite relativement (semi-)compacte dans $C_{\mathfrak{S}_i}(E_i, F)$, ce qui permet par hypothèse d'extraire de (f_n) une suite dont les restrictions à E_i convergent dans $C_{\mathfrak{S}_i}(E_i, F)$. Par le procédé diagonal, on peut alors extraire de (f_n) une suite telle que pour *tout* i, la suite des restrictions à E_i converge dans $C_{\mathfrak{S}_i}(E_i, F)$. Cette suite

converge en particulier en chacun des points de $\bigcup_i E_i$, qui est dense dans E, d'où resulte qu'elle converge dans $C_{\mathfrak{S}}(E, F)$ en vertu de nos remarques préliminaires.

Une partie du corolllaire du théorème 3 est contenu dans la proposition précédente (savoir que dans $C_{\mathfrak{S}}(E, F)$, la semi-compacité relative entraine la stricte semi-compacité relative, sous les hypothèses spécifiées dans ce corollaire). La partie la plus profonde du corollaire en question échappe pourtant à la proposition 5, en revanche nous avons le

COROLLAIRE. *Si E contient une suite partout dense, et si dans F toute partie relativement (semi-)compacte est strictement semi-compacte, alors il en est de même dans $C_{\mathfrak{S}}(E, F)$.*

Mais on notera que quelque simple que soit l'espace F (par exemple le segment compact $(0, 1)$), pour avoir des résultats dans le genre du précédent, il faut faire quelque hypothèse sur le couple $(E, \mathfrak{S})$. Ainsi, si E est un espace discret non dénombrable, et $F = (0, 1)$ on sait bien que le produit topologique $C_s(E, F)$ est compact, mais non strictement semi-compact.—Il est tout aussi évident que la moindre des choses qu'il faille supposer sur F pour avoir un résultat, c'est que dans F lui-même toute partie relativement compacte soit strictement relativement semi-compacte.

Le théorème suivant tire son intérèt du fait qu'il ne fait intervenir aucune condition de dénombrabilité sur l'espace E lui-même:

THÉORÈME 4. *Soit E un espace compact, F un espace uniforme séparé, $\mathfrak{S}$ un ensemble de parties de E recouvrant E, A une partie de $C_{\mathfrak{S}}(E, F)$ relativement semi-compacte. Supposons de plus que pour toute $f \in A$, le sous-espace $f(E)$ de F ait une topologie métrisable (il suffit donc que F ait une topologie métrisable). Alors A est strictement relativement semi-compacte.*

Soit (f_n) une suite extraite de A, tout revient à montrer qu'on peut en extraire une suite qui converge en chaque point. On est donc ramené au cas de la topologie $\mathfrak{T}_s$ de la convergence simple, et nous supposerons maintenant que F est un espace topologique séparé quelconque.—Soit B l'adhérence dans $C_s(E, F)$ de l'ensemble des f_n, considérons sur E la topologie $\mathfrak{T}'$ la moins fine rendant continues les applications éléments de B, manifestement la suite (f_n) est encore relativement semi-compacte dans l'espace $C_s(\tilde{E}, F)$, où $\tilde{E}$ désigne E muni de $\mathfrak{T}'$. Pour qu'une suite extraite de (f_n) converge en chaque point, il suffit donc qu'elle converge en chaque point d'une partie dense de $\tilde{E}$, et l'application du procédé diagonal nous ramène à montrer qu'il existe dans $\tilde{E}$ une suite partout dense. Mais $\phi(x) = \{f(x)\}_{f \in B}$ étant l'applica-

tion canonique de E dans le produit topologique $G = \prod_{f \in B} f(E)$, il revient manifestement au même de montrer que l'image $K = \phi(\tilde{E})$ admet une suite partout dense. Mais K étant compact (comme image *continue* du *compact* E) sa topologie est aussi la moins fine de celles qui rendent continues les applications f_n (topologie qui est en effet moins fine, et d'autre part *séparée* comme on vérifie aussitôt). K s'identifie donc à un sous-espace du produit topologique $\prod_n f_n(E)$, qui est *métrisable*, par conséquent K est un compact métrisable, et à fortiori séparable.

Remarque 4. Le théorème 4 vaut encore si on suppose seulement que E est semi-compact. Tout revient en effet à montrer que K est compact, mais K est déjà semi-compact comme image continue de E, d'autre part la topologique de K est complèment régulière et plus fine que la topologie métrisable définie par les f_n; la compacité de K resulte alors du théorème 3.— On aurait aussi pu s'épargner ce raisonnement et abréger en même temps la démonstration précédente en faisant usage de la proposition 1, qui dit que la suite (f_n) est encore relativement semi-compacte dans l'espace $C(\tilde{E}, F)$, lorsque $\tilde{E}$ désigne E muni de la topologie la moins fine rendant continues les f_n; tout revient alors à trouver une suite dense dans $\tilde{E}$, ce qui est immédiat.

En conjuguant le théorème 4 et la proposition 5, on obtient des cas étendus où la semi-compacité relative entraine la semi-compacité relative stricte. Le théorème de Šmulian pour la topologie faible des espaces de Banach et plus généralement des espaces ($\mathfrak{F}$) (cf. [5]) en est un cas particulier, puisque un espace ($\mathfrak{F}$) s'identifie à l'espace des formes linéaires continues sur son dual faible E', et que E' est réunion d'une suite de parties faiblement compactes. On notera d'ailleurs la parenté entre la démonstration directe du théorème de Šmulian, et celle du théorème 4. Donnons pour être complet l'énoncé le plus général du théorème de Šmulian (énoncé qui peut d'ailleurs se démontrer directement comme dans le cas classique) :

PROPOSITION 6. *Soit E un espace localement convexe, (x_n) une suite faiblement relativement semi-compacte dans E, K une partie faiblement compacte du dual E'. Alors on peut extraire de (x_n) une suite qui converge en chaque point de K (et par conséquent, en chaque point du sous-espace vectoriel faiblement fermé de E' engendré par K).—Si dans E il existe une suite de voisinages de l'origine dont l'intersection soit réduite à $\{0\}$, alors on peut extraire de (x_n) une suite faiblement convergente.*

(il suffit de noter que la dernière hypothèse assure l'existence dans E'

d'une suite de parties faiblement compactes dont la réunion soit partout dense).—Rappelons que déjà dans le dual faible d'un espace de Banach peut exister une suite relativement faiblement compacte (c'est à dire bornée), dont aucune suite extraite ne converge faiblment (cf. [5]), de sorte qu'une telle situation ne peut pas être considérée comme tératologique.

6. Critères de compacité faible dans les espaces $C^\infty(E)$. Si E est un espace topologique, nous désignons par $C(E)$ l'espace des fonctions complexes continues sur E, par $C^\infty(E)$ l'espace des fonctions complexes continues et bornées sur E, muni de la norme uniforme qui en fait un espace de Banach. Si E est compact ou semi-compact, les ensembles $C(E)$ et $C^\infty(E)$ coïncident, et nous désignerons l'espace de Banach $C^\infty(E)$ par $C(E)$ pour abréger.

Théorème 5. *Soit E un espace compact, pour qu'une partie A de $C(E)$ soit faiblement relativement compacte, il faut et il suffit qu'elle soit bornée, et relativement compacte dans $C(E)$ pour la topologie de la convergence simple.*

La nécessité de la condition est manifeste. Pour montrer qu'elle est suffisante, il suffit de montrer d'après le théorème d'Eberlein (cf. plus haut) que de toute suite (f_n) extraite de A on peut extraire une suite faiblement convergente. Mais comme une suite (g_n) extraite de (f_n) est uniformément bornée par hypothèse, et que par conséquent (les formes linéaires continues sur $C(E)$ n'étant autres que les mesures de Radon sur E) sa convergence faible équivaut à sa convergence en chaque point (théorème de Lebèsgue), il suffit donc d'extraire de la suite (f_n), relativement compacte pour la topologie de la convergence simple, une suite (g_n) qui converge en chaque point. Mais cela est possible en vertu du théorème 4.

Il faut bien noter que ce théorème n'est plus exact lorsqu'on substitue à la topologie de la convergence simple une topologie strictement moins fine, comme par exemple la topologie de la convergence en tout point sauf un seul x_0, comme on s'en convainc sans difficulté. Le théorème qui correspond au précédent et au suivant dans les espaces localement convexes généraux sera examiné en détail plus bas.

Le Théorème 5 permet l'application des critères de compacité établis au théorème 2, et notamment le critère d), qui ne fait intervenir que les valeurs des fonctions sur une partie dense de E. On a même le

Théorème 6. *Soit E un espace topologique quelconque. Pour qu'une partie A de $C^\infty(E)$ soit relativement faiblement compacte, il faut et il suffit qu'-*

elle soit bornée, et qu'il n'existe pas de suite (x_i) *extraite de* E *de suite* (f_j) *extraite de* A *telles que* $\lim_i \lim_j f_j(x_i)$ *et* $\lim_j \lim_i f_j(x_i)$ *existent et soient distincts. Ce critère subsiste si on assujettit la suite* (x_i) *à être extraite d'une partie dense* E_1 *de* E.

On sait que l'espace $C^\infty(E)$ s'identifie à l'espace des fonctions complexes continues sur la "compactification de Čech" $\hat{E}$ de E (qui s'identifie aussi à l'espace des "caractères" de l'algèbre normée complète $C^\infty(E)$—mais en fait la théorie est très élémentaire, cf. par exemple N. Bourbaki [2], page 14, exercices 6 et 7). Il existe une application canonique continue $x \to \tilde{x}$ de E sur une partie partout dense $\tilde{E}$ de $\hat{E}$ (application qui est biunivoque si et seulement si E est complètement régulier, mais peu importe), telle que l'on ait $f(x) = \tilde{f}(\tilde{x})$ quels que soient $x \in E$ et $f \in C^\infty(E)$ (où $\tilde{f}$ est la fonction sur $\hat{E}$ définie par f). D'ailleurs, l'image $\tilde{E}_1$ de E_1 dans $\hat{E}$ sera donc aussi dense. Il suffit alors d'appliquer le théorème 5 à l'espace $C(\hat{E})$, puis le corollaire 2 du théorème 2 à ce même espace et la partie dense $\tilde{E}_1$ de $\hat{E}$.— Notons que l'application de ce dernier théorème et du critère du corollaire 2 du théorème 2, montre aussitôt que le théorème 5 reste valable si E est seulement semi-compact.

Donnons une application immédiate du théorème 6. Si G est un semi-groupe, muni éventuellement d'une topologie qui rende continues ses translations à gauche et à droite, nous dirons avec F. Eberlein ([6]) qu'une fonction complexe bornée et continue sur G est faiblement presque-périodique à gauche (resp. à droite), si l'ensemble de ses translatées gauches (respectivement droites) est une partie relativement faiblement compacte de l'espace de Banach $C^\infty(G)$. On a alors immédiatement la

PROPOSITION 7. *Pour qu'une* $f \in C^\infty(G)$ *soit faiblement presque-périodique à gauche (ou à droite) il faut et il suffit qu'il n'existe pas de suites* (x_i) *et* (y_j) *extraites de* G, *telles que* $\lim_i \lim_j f(x_i y_j)$ *et* $\lim_j \lim_i f(x_i y_j)$ *existent et soient distincts. En particulier, les fonctions faiblement presque périodiques à gauche et à droite sont les mêmes.* Il sera donc à propos de les appeller fonctions faiblement presque-périodiques tout court).

7. Renforcement des critères de faible compacité relative dans les espaces vectoriels localement convexes.

THÉORÈME 7. *Soit* E *un espace vectoriel localement convexe séparé,* (K_α) *une famille de parties du dual* E' *de* E, *à enveloppes convexes cerclées*

relativement faiblement compactes, et telles que la famille des enveloppes convexes cerclées fermées $\widetilde{K}_\alpha$ des K_α engendre algébriquement tout E'. Soit A une partie bornée de E, et supposons E complet et pour la topologie $\mathfrak{T}$ de la convergence uniforme sur les K_α, ou du moins l'enveloppe convexe fermée de A complète pour cette topologie.

a) *Si les K_α sont faiblement fermés (c'est à dire faiblement compacts) alors pour que A soit faiblement relativement compact dans E, il faut et il suffit que pour tout α, l'ensemble des fonctions continues sur K_α définies par les éléments de A soit relativement compact dans $C(K_\alpha)$ pour la topologie de la convergence simple.*

b) *Si on ne suppose plus forcément les K_α fermés, une condition nécessaire et suffisante pour que A soit relativement faiblement compact, est qu'il n'existe pas de suite (x_i) extraite de A et de suite (x'_j) extraite de quelque K_α, telles que* $\lim_i . \lim_j . \langle x_i, x'_j\rangle$ *et* $\lim_j . \lim_i . \langle x_i, x'_j\rangle$ *existent et soient distincts.*

En vertu du théorème 2 corollaire 2 (qui s'applique ici puisque A est borné), la condition énoncée dans b) équivaut à la condition énoncée dans a), appliquée aux adhérences faibles des K_α, de sorte qu'on peut se borner à démontrer a). Nous identifions comme d'habitude E à l'espaces des formes linéaires continues sur son dual faible, et notons comme dans 4. que tout revient à montrer que pour toute forme linéaire X sur E' qui est faiblement adhérente à A, les restrictions aux $\widetilde{K}_\alpha$ sont faiblement continues. (Dans la suite, il est inutile de conserver l'indice α). D'après le théorème de Mackey ([9]), le dual de E muni de $\mathfrak{T}$ est encore E'. Il existe d'autre part une application linéaire canonique $x \to u(x)$ de E dans l'espace de Banach $C(K)$ des fonction complexes continues sur K, application qui est continue par la définition même de $\mathfrak{T}$, et dont la transposée u' est donc une application faiblement continue du dual C' de $C = C(K)$ dans E'. L'image de la boule unité B de C' par u' est donc une partie convexe cerclée faiblement compacte de E' (puisque B est faiblement compacte), contenant évidemment K, donc aussi $\widetilde{K}$. En fait, il nous sera commode de savoir qu'elle est même identique à $\widetilde{K}$, cela resulte immédiatement du fait connu que B est l'enveloppe convexe cerclée faiblement fermée dans C' de l'ensemble des " masses $+1$ " placées aux divers points de K (comme il resulte aussitôt de l'emploi des ensembles polaires, cf. [5]). Nous allons montrer que la restriction de X à $\widetilde{K}$ est de la forme $\langle X, u'.\mu\rangle = \langle f, \mu\rangle$, μ désignant l'élément générique de B, et où f est un élément convenable de l'espace $C = C(K)$ (c'est en fait la fonction sur $K: f(x') = \langle X, x'\rangle$), il s'ensuivra aussitôt que la restriction de X à $\widetilde{K}$ est

continue, puisqu'en la composant avec l'application continue u' du compact B *sur* $\tilde{K}$, on obtient une application continue (cf. [1] page 53, th. 1, et page 62 th. 1, cor. 2). Soit donc ϕ la trace sur A du filtre des voinages faibles de X, on a pour tout $x' = u'.\mu(\mu \varepsilon B)$:

$$\langle X, u'.\mu\rangle = \lim_{\phi}. \langle x, u'.\mu\rangle = \lim_{\phi}. \langle u.x, \mu\rangle$$

or l'image de ϕ par u est un filtre de Cauchy pour la convergence simple, et $u(A)$ est *faiblement* relativement compact dans $C(K)$, comme il resulte de l'hypothèse et au théorème 5; il suit que $u.x$ tend *faiblement* suivant ϕ vers une limite $f \varepsilon C(K)$, d'où suit bien $\langle X, u'.\mu\rangle = \langle f, \mu\rangle$.

COROLLAIRE 1. *Soit E un espace de Banach, K l'ensemble des points extrémaux de la boule unité de son dual. Pour que $A \subset E$ soit faiblement relativement compact, il faut et il suffit qu'il n'existe pas de suite (x_i) extraite de A et de suite (x'_j) extraite de K, telle que $\lim_i.\lim_j.\langle x_i, x'_j\rangle$ et $\lim_j.\lim_i.\langle x_i, x'_j\rangle$ existent et soient distincts.*

COROLLAIRE 2. *Soit E un espace localement convexe, (B_α) une famille de parties bornées de E telle que toute partie bornée de E soit contenue dans l'enveloppe convexe cerclée fermée de quelque B_α. Supposons le dual fort* (cf. [5]) *E' de E complet, ou du moins ses parties bornées et fermées complètes. Pour que $A \subset E'$ soit relativement compact pour la topologie $\sigma(E', E'')$ (E'' désignant le dual de E' fort) il faut et il suffit qu'elle soit fortement bornée, et qu'il n'existe pas de suite (x_i) extraite de quelques B_α et de suite (x'_j) extraite de A, telles que $\lim_i.\lim_j.\langle x_i, x'_j\rangle$ et $\lim_j.\lim_i.\langle x_i, x'_j\rangle$ existent et soient distincts.*

Remarque 5. En fait, sous les conditions du théorème 7, on peut même affirmer que l'enveloppe convexe fermée de A est faiblement compacte. En effet, la démonstration d'un théorème connu de Krein pour les espaces de Banach se transpose aux espace vectoriels localement convexes pour donner l'énoncé suivant: Soit E un espace localement convexe séparé, A une partie faiblement relativement compacte; pour que son enveloppe convexe fermée soit faiblement compacte, il faut et il suffit qu'elle soit complète (ne fût-ce d'ailleurs que pour la topologie $\tau(E, E')$ associée).—Nous ne donnerons pas la démonstration de cette proposition, qui s'appuie essentiellement sur le théorème d'Eberlein généralisé (proposition 2) et le resultat de [7] rappelé plus haut qui nous a déjà servi pour la proposition 2 et le théorème 7.

Notons encore que le théorème 6 donne comme corollaire immédiat un

critère de convergence faible d'une suite dans un espace $C^{\infty}(E)$, et le théorème 7 un critère de convergence faible d'un suite dans un espace de Banach quelconque. Ces critères, pour le cas particulier de suites tendant faiblement vers 0, se trouvent déjà dans Banach ([4], page 222). Il ne semble pas possible d'ailleurs d'en déduire les théorèmes 6 et 7.

INSTITUT DE MATHÉMATIQUES, NANCY (M. ET M.), FRANCE.

BIBLIOGRAPHIE.

[1] N. Bourbaki, Topologie Générale, Chap. 1 et 2, *Actualités Scient. et Ind.*, No. 858 (1940).

[2] ———, Topologie Générale, Chap. 9, *ibid.*, No. 1045.

[3] ———, Topologie Générale, Chap. 10, *ibid.*, No. 1084.

[4] S. Banach, Théorie des opérations linéaires, Varsovie, 1932.

[5] J. Dieudonné et L. Schwartz, "La dualité dans les espaces ($\mathfrak{F}$) et ($\mathfrak{LF}$)," *Annales de l'Institut Fourier* (Grenoble), vol. 1 (1950), pp. 61-101.

[6] F. Eberlein, "Abstract ergodic theorems and weak almost periodic functions," *Transactions of the American Mathematical Society*, vol. 67 (1949), pp. 217-240.

[7] A. Grothendieck, "Sur la complétion du dual d'un espace localement convexe," *Comptes Rendus*, vol. 230, pp. 605-606.

[8] G. Köthe, "Erweiterung von Linearfunktionen in linearen Räumen," *Mathematische Annalen*, vol. 116 (1939), pp. 719-732.

[9] G. W. Mackey, "On convex topological linear spaces," *Transactions of the American Mathematical Society*, vol. 60 (1946), pp. 520-537.

[10] A. Weil, "Sur les espaces à structure uniforme et la Topologie Générale," *Actualités Scient. et Ind.*, No. 551 (1937).

GROUPS AND CARDINAL NUMBERS.*

By W. R. Scott.

1. Introduction. Let G be an infinite group. For $x \varepsilon G$, let $E(x, G) = E(x)$ be the set of $g \varepsilon G$ such that the equation $g^n = x$ has no solution for n. Let $K = K(G)$ be the set of $k \varepsilon G$ such that $o(E(k)) < o(G)$, where $o(S)$ is the cardinal number of elements in S. Observe that if G is a p^∞ group, then $K(G) = G$.

This paper is devoted primarily to the problem of determining K, given G. The problem is solved for Abelian groups (Corollary 1 to Theorem 8, and Theorems 9 and 10), and considerable progress is made in the general case (Theorems 1-8, Corollary 2 to Theorem 8, and the Corollary to Theorem 11).

In the process of (partial) solution of the above problem, two sidepaths are investigated. In section **3**, a few results are given concerning the size of layers of a group. These results are perhaps of independent interest. In sections **4**, **5**, and **6**, in conjunction with the study of K, the intersection D of all subgroups G_α of G with $o(G_\alpha) = o(G)$ is introduced. The principal result is that $D = K$ for Abelian groups G.

2. Definitions and notations. Let G be a group written multiplicatively with identity e. If $g \varepsilon G$, $o(g)$ will denote the order of g; $o(g) = \infty$ means that $g^n = e$ implies $n = 0$. G is *periodic* if $o(g) < \infty$ for every $g \varepsilon G$. G is *locally cyclic* if for every $g_1, g_2 \varepsilon G$ there exists a $g \varepsilon G$ and integers m and n such that $g^m = g_1$, $g^n = g_2$. A subgroup H of G is *central* if $H \subseteq Z$ where Z is the center of G. A subgroup H is *strictly characteristic* if $\sigma(H) \subseteq H$ for every endomorphism $\sigma(G) = G$ of G onto G. It is *fully characteristic* if $\sigma(H) \subseteq H$ for every endomorphism σ of G. The letter p will always denote a prime. A *p-group* is a group G such that $o(g) = p^{r(g)}$ for all $g \varepsilon G$, for a fixed prime p. A *layer* $L(n)$ $(L(\infty))$ of G is the set of $g \varepsilon G$ with $o(g) = n(\infty)$. The notations $o(S)$, $E(x)$, K, and $D(G) = D$ will be used as in section **1**. If G is Abelian, the *torsion* T of G is the subgroup of $g \varepsilon G$ with $o(g) < \infty$. The *p-component* T_p of T is defined by the equation $T_p = \bigcup_{i=0}^{\infty} L(p^i)$. For Abelian groups G, let $H(p, r) = \bigcup_{i=0}^{r} L(p^i)$. A p^∞

* Received January 15, 1951; revised October 9, 1951.

group is a p-component of the group of rationals mod 1. A direct product of subgroups of a group will be denoted by Π. *The symbol* $\bigcup$ *will always denote the point set union.* If S_1 and S_2 are sets, $S_1 - S_2$ will denote the set of s such that $s \in S_1$, $s \notin S_2$. For the sake of brevity, a cardinal A will be called *standard* if (i) $A > \aleph_0$ and (ii) A is not the sum of $\aleph_0$ smaller cardinals. In the statement of a theorem, **H.** and **C.** will mean hypothesis and conclusion respectively.

3. The layers. Lemma 1. **H.** *G is an infinite group, P a set of primes.* $S = (\bigcup_{\lambda} \bigcup_{p \notin P} L(\lambda p)) \cup L(\infty)$.

C. $o(S) = o(G)$ *or* 0.

Proof. Suppose that there exists an $s \in S$. Clearly, we may suppose that either $o(s) = \infty$ or $o(s) = p \notin P$. Let N be the normalizer of s. Then $i(N) = o(Cl(s))$, where $i(N)$ denotes the index of N and $Cl(s)$ the class of conjugates of s. If $g \in Cl(s)$, then $o(g) = o(s)$, and therefore $Cl(s) \subseteq S$. Hence, if $i(N) = o(G)$, we have $o(S) = o(G)$. Therefore suppose $i(N) < o(G)$. Since $o(G) = o(N)i(N)$, and since $o(G) \geqq \aleph_0$, we have $o(N) = o(G)$. If $o(N \cap S) = o(G)$, we are done. If not, then $o(N \cap (G - S) = o(G)$. Let $g \in N \cap (G - S)$. Then $(gs)^r = e$, $r \neq 0$, implies $g^r s^r = e$, which implies that s^r is in the subgroup $\{g\}$ generated by g. If $o(s) = \infty$. this would imply that $o(g) = \infty$, a contradiction since $g \notin S$. If $o(s) = p \notin P$, then $p | r$, whence $p | o(gs)$. Thus $gs \in S$ for all $g \in N \cap (G - S)$. Hence

$$o(S) \geqq o(N \cap (G - S)) = o(G),$$

i. e. $o(S) = o(G)$, and the theorem is proved.

Corollary 1. **H.** *G is a non-periodic group.*

C. $o(L(\infty)) = o(G)$.

Proof. Since G is non-periodic it is infinite. In Lemma 1, let P be the set of all primes.

Corollary 2. **H.** *p is a fixed prime. G is an infinite periodic group which is not a p-group.* $S = \bigcup_{n \neq p^r} L(n)$.

C. $o(S) = o(G)$.

Lemma 2. **H.** $n = \prod_{i=1}^{r} p_i^{\alpha_i}$, *with $p_i \neq p_j$ for $i \neq j$, and $\alpha_i > 0$. G is a group (finite or infinite).*

C. $o(L(n)) \leqq \prod_{i=1}^{r} o(L(p_i^{\alpha_i}))$.

Proof. If $g \in L(n)$, then $g = \prod_{i=1}^{r} g_i$ where $g_i \in L(p_i^{\alpha_i})$.

LEMMA 3. **H.** *G is an Abelian p-group.*

C. $o(L(p^r)) \leqq o(L(p^{r-1}))(o(L(p)) + 1) \leqq (o(L(p)) + 1)^r$, $r = 1, 2, \cdots$.

Proof. If $g_1^p = g_2^p$, then $(g_1 g_2^{-1})^p = e$, and conversely. Hence, if $g \in G$, there are 0 or $(o(L(p)) + 1)$ solutions of the equation $x^p = g$. Therefore there are at most $o(L(p^{r-1}))(o(L(p)) + 1)$ elements $x \in G$ such that $x^p \in L(p^{r-1})$; i. e. $o(L(p^r)) \leqq o(L(p^{r-1}))(o(L(p)) + 1)$. The second inequality follows from the first by induction.

COROLLARY 1. **H.** *G is an Abelian p-group.* $o(G) > \aleph_0$. $H(p, 1) = L(1) \cup L(p)$.

C. $o(H) = o(G)$.

Proof. If H is finite, then by Lemma 3, G is denumerable, a contradiction. Hence, again by Lemma 3,

$$o(G) = \sum_{i=0}^{\infty} o(L(p^i)) = o(L(p)) = o(H).$$

COROLLARY 2. **H.** *G is Abelian and periodic; $o(G)$ is standard.*

C. *There exist a p such that the subgroup $H(p, 1)$ satisfies the relation*

$$o(H(p, 1)) = o(G).$$

Proof. $G = \Pi G_p$ where G_p is the p-component of G. Hence

$$o(G) \leqq \sum_{r=1}^{\infty} \prod_{i=1}^{r} o(G_{p_i}).$$

Since $o(G)$ is standard, this implies that $o(G_p) = o(G) > \aleph_0$ for some p. The theorem follows from Corollary 1.

Remark. The condition that $o(G)$ be standard cannot be omitted (for infinite groups) in this corollary.

4. Formulas and lemmas. We first list a few elementary identities and inclusion relations for $E(x)$.

(1) $E(e)$ is the empty set.

(2) $E(x^{-1}) = E(x)$.

For $g^n = x$ if and only if $g^{-n} = x^{-1}$.

(3) $E(x_1x_2) \subseteq E(x_1) \cup E(x_2)$.

For $g^m = x_1$, $g^n = x_2$ imply $g^{m+n} = x_1x_2$.

(4) $E(\sigma(x), \sigma(G)) \subseteq \sigma(E(x, G))$ for any endomorphism σ of G.

For $g^n = x$ implies $(\sigma(g))^n = \sigma(x)$.

(5) $E(h, H) = H \cap E(h, G)$ if $h \varepsilon H \subset G$.

(6) $E(x) \supseteq G - N(x)$, where $N(x)$ is the normalizer of x.

For if $g^n = x$, then $gx = xg$, and $g \varepsilon N(x)$.

THEOREM 1. **H.** *G is an infinite group.*

C. *D is a strictly characteristic subgroup of G. Moreover $K \subseteq D$.*

Proof. Let $d \varepsilon D$, let $\sigma(G) = G$ be an endomorphism of G onto G, and let G_α be a subgroup of G with $o(G_\alpha) = o(G)$. Then if $H = \sigma^{-1}(G_\alpha)$, H is a subgroup of G with $o(H) = o(G)$. Thus $d \varepsilon H$, and therefore $\sigma(d) \varepsilon G_\alpha$. Therefore $\sigma(d) \varepsilon \cap G_\alpha = D$. Hence D is strictly characteristic.

If there exists a $k \varepsilon K - D$, then $k \not\varepsilon G_\alpha$ for some subgroup G_α with $o(G_\alpha) = o(G)$. It follows easily that $G_\alpha \subseteq E(k)$. This gives $o(E(k)) = o(G)$ which is a contradiction. Thus $K \subseteq D$.

COROLLARY. **H.** *G is an infinite group such that $G = \Pi G_\alpha$ with $o(G_\alpha) < o(G)$.*

C. $K = D = e$.

Proof. Let $G_\alpha^* = \prod_{\beta \neq \alpha} G_\beta$. Then $G = G_\alpha \times G_\alpha^*$, and therefore $o(G_\alpha^*) = o(G)$. Hence by Theorem 1, $K \subseteq D \subseteq \cap G_\alpha^* = e$.

LEMMA 4. **H.** *G_n is an infinite group, $n = 1, 2, \cdots$; $G_1 \subseteq G_2 \subseteq \cdots$; $G = \cup G_n$.*

C. $K(G) \subseteq \operatorname{Lim\,inf} K(G_n)$.

Proof. Let $k \varepsilon K(G)$. Then $o(E(k)) < o(G) = \lim o(G_n)$. Hence there exists an $n_0(k)$ such that if $n > n_0$, then (i) $k \varepsilon G_n$, (ii) $o(E(k, G)) < o(G_n)$. Hence by (5), $o(E(k, G_n)) \leqq o(E(k, G)) < o(G_n)$ and $k \varepsilon K(G_n)$ if $n > n_0$, i. e. $k \varepsilon \liminf K(G_n)$.

Remark. If $o(G)$ is standard, then $K(G) = \lim K(G_n)$ and $D(G) = \lim D(G_n)$.

LEMMA 5. **H.** *H is a subgroup of G. $o(G) = o(H) \geqq \aleph_0$.*

C. $K(G) \subseteq K(H)$,

$D(G) \subseteq D(H)$.

Proof. By Theorem 1, $K(G) \subseteq H$. By (5), if $k \varepsilon K(G)$, then $o(E(k, H)) \leqq o(E(k, G)) < o(G) = o(H)$. Hence $k \varepsilon K(H)$. The other half of the theorem follows from the definition of D.

LEMMA 6. **H.** *$o(G) \geqq \aleph_0$. H is a proper subgroup of G.*

C. $o(G - H) = o(G)$.

Proof. If $o(H) < o(G)$, this is immediate. If $o(H) = o(G)$. then for any $g \varepsilon G - H$, $Hg \subseteq G - H$ and we have

$$o(G) \geqq o(G - H) \geqq o(Hg) = o(H) = o(G).$$

5. The subgroup K, general case.

THEOREM 2. **H.** *G is an infinite group.*

C. *K is a central, locally cyclic, periodic subgroup of G.*

Proof. (i) K is a subgroup of G. By (1), $e \varepsilon K$. By (2), if $k \varepsilon K$, then $k^{-1} \varepsilon K$. By (3), if $k_1, k_2 \varepsilon K$, then $o(E(k_1 k_2)) \leqq o(E(k_1)) + o(E(k_2)) < o(G)$, and $k_1 k_2 \varepsilon K$.

(ii) K is central. For let $k \varepsilon K$, and let N be the normalizer of k in G. Then by (6) and Lemma 6, if $N \neq G$, we get $o(E(k)) \geqq o(G - N) = o(G)$, which is a contradiction. Hence $N = G$, i. e. $k \varepsilon Z$. Thus $K \subseteq Z$ and K is central.

(iii) K is locally cyclic. Let $k_1, k_2 \varepsilon K$, $k_i \neq e$ (the other case is trivial). Then there exists an $x \varepsilon G - (E(k_1) \cup E(k_2))$. Hence there exists integers n_1, n_2 such that $x^{n_1} = k_1$, $x^{n_2} = k_2$. Let $d = (n_1, n_2)$, and let $d = r_1 n_1 + r_2 n_2$. Hence $x^d = k_1^{r_1} k_2^{r_2} \varepsilon K$. Moreover $k_i = (x^d)^{n_i/d}$. Thus K is locally cyclic.

(iv) K is periodic. Suppose, on the contrary, that there exists a $k \varepsilon K$ with $o(k) = \infty$. Then if $x \not\varepsilon E(k)$, there exists an $n(x)$ such that $x^{n(x)} = k$. Thus $o(x) = \infty$ also. Let S^+ be the set of $x \not\varepsilon E(k)$ for which (the unique) $n(x) > 1$, and S^- the set of $x \not\varepsilon E(k)$ for which $n(x) < -1$. Thus

$$G = E(k) \cup k \cup k^{-1} \cup S^+ \cup S^-,$$

and therefore either $o(S^+) = o(G)$ or $o(S^-) = o(G)$, or both hold. If $x \varepsilon S^+$, and r is any integer, then

$$(kx)^r = x^{r(n(x)+1)} \neq x^{n(x)} = k.$$

Thus $kx \varepsilon E(k)$, and therefore $kS^+ \subseteq E(k)$. Similarly $k^{-1}S^- \subseteq E(k)$. Since $o(kS^+) = o(S^+)$, $o(k^{-1}S^-) = o(S^-)$, this leads to a contradiction.

THEOREM 3. **H.** *G is not periodic.*

C. $K = e$.

Proof. By Corollary 1 of Lemma 1, $o(L(\infty)) = o(G)$. If $k \varepsilon K$, $k \neq e$, then $o(k) < \infty$ by Theorem 2. Hence $L(\infty) \subseteq E(k)$, and $o(E(k)) = o(G)$, a contradiction. Therefore $K = e$.

THEOREM 4. **H.** *$o(G)$ is standard.*

C. *K is cyclic of order p^n for some prime p and some integers n.*

Proof. By Theorem 3, we may assume that G is periodic. Therefore $o(G) = \Sigma o(L(n))$. Since $o(G)$ is standard, $o(L(n)) = o(G) > \aleph_0$ for some integer n. Therefore by Lemma 2, $o(L(p^r)) = o(G)$ for some prime p and some positive integer r. If $x \not\in \bigcup_{i=0}^{r} L(p^i)$, then $L(p^r) \subseteq E(x)$ and $x \not\in K$. Thus $K \subseteq \bigcup_{i=0}^{r} L(p^i)$ (actually $K \subseteq \bigcup_{i=0}^{r-1} L(p^i)$). Now since K is locally cyclic and periodic, it is isomorphic to a subgroup of the group R_1 of rationals mod 1. Hence K is cyclic of order p^n, $0 \leqq n \leqq r$ (actually $n < r$).

THEOREM 5. **H.** *G is an infinite group.*

C. *K is either a cyclic group of order p^n or a p^∞ group.*

Proof. Case 1. $o(G) > \aleph_0$. If $o(G)$ is standard, the conclusion is immediate from Theorem 4. If $o(G)$ is not standard, then there exists an increasing sequence $\{B_n\}$ of standard cardinals whose sum is $o(G)$. Choose a sequence $\{G_n\}$ of subgroups such that (α) $G_n \subseteq G_{n+1}$, (β) $o(G_n) = B_n$, and (γ) $G = \cup G_n$. Then by Theorem 4, $K(G_n)$ is cyclic of order $p_n^{t_n}$. It follows readily that $\liminf K(G_n)$ is a subgroup of G and is one of the two types described in the theorem. By Lemma 4, $K(G)$ is also of the required type.

Case 2. $o(G) = \aleph_0$. By Theorem 3, we may assume that G is periodic. If $o(L(p)) > 0$ for an infinity of primes p, then for any $x \neq e$ we have $o(E(x)) = \aleph_0$, and therefore $K = e$. If for some p, $S_p = \bigcup_{r=0}^{\infty} L(p^r)$ is infinite, then it easily follows that K is a p-group. But a locally cyclic p-group is one of the two types described. If, finally, S_p is finite for all p, and $o(S_p) = 1$ for all but a finite number of p, say $p_1, \cdots, p_r$, then there is a maximum β_i

such that there exist $g \varepsilon S_{p_i}$ with $o(g) = p_i^{\beta_i}$. Then no element of G has order greater than $n = \Pi p_i^{\beta_i}$. Hence $o(G) = \sum_{j=1}^{n} o(L(j))$ which is finite by Lemma 2. This is a contradiction. Hence the theorem is true in any case.

Note that Theorem 5 improves Theorem 2.

THEOREM 6. **H.** *G is an infinite group.*

C. *K is a fully characteristic subgroup of G.*

Proof. Let σ be any endomorphism of G and let N be the kernel of σ. If $o(N) = o(G)$, then $K \subseteq D \subseteq N$, and $\sigma(K) = e \varepsilon K$. If $o(N) < o(G)$, then $\sigma(G)$ is a subgroup of G with $o(\sigma(G)) = o(G)$. If $k \varepsilon K$, then by (4)

$$o(E(\sigma(k), \sigma(G))) \leqq o(\sigma(E(k, G))) \leqq o(E(k, G))$$
$$< o(G) = o(\sigma(G)).$$

Hence $\sigma(K) \subseteq K(\sigma(G))$. Now by Lemma 5, $K(G) \subseteq K(\sigma(G))$, and by Theorem 5, $K(\sigma(G))$ is either a cyclic group of order p^n or a p^∞ group. Since $o(\sigma(K)) \leqq o(K)$, and since both K and $\sigma(K)$ are subgroups of $K(\sigma(G))$, it follows that $\sigma(K) \subseteq K$. Hence K is fully characteristic.

THEOREM 7. **H.** *G is an infinite group. $G = H \times F$ where H is a p-group and F is a periodic subgroup of G such that* (i) $o(F) < o(G)$ *and* (ii) $f \varepsilon F$ *implies* $p \nmid o(f)$.

C. $K(G) = K(H)$.

Proof. By (i) $o(H) = o(G)$, whence by Lemma 5, $K(G) \subseteq K(H)$. Let $k \varepsilon K(H)$. Then if $h \varepsilon H - E(k, H)$, we have $h^n = k$ for some n. Let $o(h) = p^s$ and let $f \varepsilon F$, $o(f) = r$. Then there exists an integer λ such that $\lambda r \equiv n \pmod{p^s}$. Thus $(hf)^{\lambda r} = h^{\lambda r} = h^n = k$. Therefore $hf \varepsilon G - E(k, G)$. Hence

$$o(E(k, G)) \leqq o(E(k, H))o(F) < o(G).$$

Thus $k \varepsilon K(G)$. This proves that $K(G) = K(H)$.

THEOREM 8. **C.** *K is a p^∞ group if and only if there exists a central p^∞ subgroup C such that G/C is finite. If such a C exists, then $C = K = D$.*

Proof. Suppose first that such a C exists. Then G is periodic. Let $g_1, \cdots, g_n$ be representatives of the cosets of C. Then there exists an $r > 0$ such that $g_i^r = e$, $i = 1, \cdots, n$. Choose t such that $p^t \geqq r$. Then if $c \varepsilon C$, $o(c) = p^s$, and $c' \varepsilon C$, $o(c') \geqq p^{s+t}$, we have $(c'g_i)^r = c'^r$ which has order

$\geqq p^s$. Hence a suitable power of $c'g_i$ equals c. Thus $E(c)$ is finite, and $c \in K$. Hence $C \subseteq K$, and by Theorem 5, $C = K$. Also, clearly $K = D$.

Conversely suppose that K is a p^∞ group. K is central by Theorem 2. If G/K is finite, we are done. Suppose that G/K is infinite. Then $o(G/K) = o(G)$. Now let $g \in G$, and let n be the smallest positive integer such that $g^n \in K$. Then $g^n = k$ and $k'^n = k^{-1}$ for some k, $k' \in K$. Thus $(gk')^n = e$ while $(gk')^r \notin K$ for $0 < r < n$. Therefore if $k'' \in K$, $k'' \neq e$, then $gk' \in E(k'')$, i. e. $E(k'')$ contains at least one element of each coset of K. Thus $o(E(k'')) \geqq o(G/K) = o(G)$, a contradiction. Therefore G/K is finite as asserted.

Corollary 1. **H.** *G is abelian.*

C. *K is a p^∞ group if and only if $G = H \times F$ where H is a p^∞ group and F is finite.*

Proof. If K is a p^∞ group then (see [1], p. 767) $G = K \times F$, and by Theorem 8, F is finite. The converse is obvious.

Corollary 2. **H.** $o(G) > \aleph_0$.

C. *K is a cyclic group of order p^n.*

Proof. This follows from Theorems 5 and 8.

Corollary 3. **H.** $K_2 = K(G/K(G))$.

C. *K_2 has order* 1 *or is not defined.*

Proof. If K is a p^∞ group, then G/K is finite and K_2 is not defined. Otherwise K is finite. Let $Kx \in K_2$. Then if E is the set of $g \in G$ such that $Kg \in E(Kx)$, we have

$$o(E) = o(K)o(E(Kx)) < o(G).$$

If $g \notin E \cup (\cup E(k))$, then $g^n = kx$ for some integer n and some $k \in K$. and $g^m = k^{-1}$ for some m. Hence $g^{m+n} = x$, and $g \notin E(x)$. Therefore

$$o(E(x)) \leqq o(E) + \sum_{k \in K} o(E(k)) < o(G)$$

since K is finite. Hence $x \in K$ and K_2 has order 1.

6. Abelian groups.

Lemma 7. **H.** $G \supseteq \Pi H_\alpha$, $\alpha \in S$, *where* $o(G) = o(S) \geqq \aleph_0$. *$R$ is the set of subgroups G_β of G with $o(G_\beta) = o(G)$.*

C. (i) $D = e$. (ii) $o(R) = 2^{o(G)}$.

Proof. There are $2^{o(G)}$ subset of G, hence $o(R) \leqq 2^{o(G)}$. There are $2^{o(G)}$ subsets S' of S of order $o(G)$. Then if $N(S') = \Pi H_\alpha$, $\alpha \varepsilon S'$, we have $o(N(S')) = o(G)$, and $S' \neq S''$ implies $N(S') \neq N(S'')$. Hence $o(R) = 2^{o(G)}$. Clearly, also, $D \subseteq \cap N(S') = e$.

THEOREM 9. **H.** *$o(G) > \aleph_0$, and G is Abelian. R is the set of subgroups G_α of G with $o(G_\alpha) = o(G)$.*

C. (i) $o(R) = 2^{o(G)}$. (ii) $K = D = e$.

Proof. Let T be the torsion of G.

Case 1. $o(T) < o(G)$. Then $o(L(\infty)) = o(G)$. Let $B = \{b_\alpha\}$ be a maximal set of independent elements of $L(\infty)$. We assert that $o(B) = o(L(\infty))$. For suppose that $o(B) < o(L(\infty))$ and let H be the subgroup of G generated by the b_α. Then $o(H) = o(B)$ if $o(B) \geqq \aleph_0$, and $o(H) = \aleph_0$ if $o(B) < \aleph_0$. Hence, in any case, $o(H) < o(L(\infty))$. For fixed n, if $x^n = y^n$, then $(xy^{-1})^n = e$ and $xy^{-1} \varepsilon T$. Thus there are at most $o(T)$ solutions of $x^n = h$ for fixed n and fixed $h \varepsilon H$. Hence the number of solutions of $x^n \varepsilon H$ for x, allowing n to vary, $n \neq 0$, is at most $\aleph_0 o(T) o(H) < o(G) = o(L(\infty))$. Therefore there exists an $x \varepsilon L(\infty)$-H such that the set $B' = B \cup x$ is independent. This contradicts the maximality of B. Hence $o(B) = o(L(\infty))$ as asserted. The theorem follows in this case from Lemma 7.

Case 2. $o(T) = o(G)$. Then $T = \Pi T_p$.

Case 2.1. $o(T_p) = o(T)$ *for* some p. Then by Corollary 1 of Lemma 3, we have $o(H(p,1)) = o(G)$. Now $H(p,1) = \Pi C_\alpha$ where the C_α are cyclic of order p. There are clearly $o(G)$ factors C_α. Therefore by Lemma 7 the theorem is true.

Case 2.2. $o(T_p) < o(T)$ for all p. Let $U = \Pi T_{p_i}$ for all primes p_i such that $o(T_{p_i}) > \aleph_0$. Then $o(U) = o(T) = \Sigma o(T_{p_i})$. By Case 2.1 each T_{p_i} has $2^{o(T_{p_i})}$ subgroups $H(i)$ of order $o(T_{p_i})$. For each i, choose an $H(i) \subseteq T_{p_i}$, with $o(H(i)) = o(T_{p_i})$. Then $V = \Pi H(i)$ is a subgroup of U such that $o(V) = o(U)$. The number of subgroups formed in this manner is clearly

$$\Pi 2^{o(T_{p_i})} = 2^{\Sigma o(T_{p_i})} = 2^{o(U)} = 2^{o(G)}.$$

Moreover, it is clear from Case 2.1 that the intersection of all the subgroups V is e.

Remark 1. This result may be combined with that in [3] to get the following theorem:

If G is an Abelian group such that for any proper subgroup H it is true that $o(H) < o(G)$, then G is either a finite group or a p^∞ group.

Remark 2. The proof of Theorem 9 may be altered slightly to prove the following generalization:

If G is a non-denumerable Abelian group, $o(G) \geqq A \geqq \aleph_0$, and $R(A)$ is the set of subgroups G_α of G with $o(G_\alpha) = A$, then $o(R) = (o(G))^A$.

THEOREM 10. **H.** *G is Abelian, $o(G) = \aleph_0$, and G cannot be expressed in the form $H \times F$ with H a p^∞ group and F finite.*

C. $D = K = e$.

Proof. Case 1. G is not periodic, i. e., there exists a $g \varepsilon G$ with $o(g) = \infty$. If $g' \varepsilon G$, $1 < o(g') < \infty$, then g' is not in the subgroup generated by g. If $g'' \varepsilon G$, $o(g'') = \infty$, then g'' is not in the subgroup generated by g''^2. Hence $D = e$.

Case 2. G is periodic. Then $G = \Pi G_p$.

Case 2.1. $o(G_p) = o(G)$ for some p.

Case 2.1.1. $o(H(p,r)) = o(G)$ for some r. Then $H(p,r) = \prod_{n=1}^{\infty} C_n$ where C_n is a cyclic group. Then by Lemma 7, $D = e$.

Case 2.1.2. $o(H(p,r)) < o(G)$ for all r, i. e., $H(p,r)$ is finite for all r. Then (see [2], p. 102) there exists a p^∞ subgroup H of G. Hence (see [1], p. 767) $G = H \times M$. By hypothesis M is infinite, and therefore $D \subseteq H \cap M = e$.

Case 2.2. $o(G_p) < o(G)$ for all p. Then by the Corollary to Theorem 1, $D = e$.

Thus in all cases $D = e$, and therefore $K = e$ also.

7. Miscellaneous.

THEOREM 11. **H.** *N is a normal subgroup of G, $o(G/N) \geqq \aleph_0$, $D(G/N) = eN$.*

C. $D(G) \subseteq N$.

Proof. If $o(N) = o(G)$, the conclusion follows from the definition of

$D(G)$. Suppose $o(N) < o(G)$, and let $d \varepsilon D(G)$. If H^* is a subgroup of G/N not containing dN, then its inverse image H is a subgroup of G not containing d. Hence $o(G/N) = o(G) > o(H) \geqq o(H^*)$. Thus $dN \varepsilon D(G/N)$, i. e. $dN = eN$. Hence $d \varepsilon N$.

Remark. The above theorem remains true if the letter D is replaced by K throughout.

COROLLARY. **H.** $o(G) > \aleph_0$. *Q is the commutator subgroup of G.* **C.** *$D \subseteq Q$, hence $K \subseteq Q$.*

Proof. If $o(Q) = o(G)$ then $D \subseteq Q$. If $o(Q) < o(G)$, then $o(G/Q) = o(G) > \aleph_0$, and G/Q is Abelian. By Theorem 9, $D(G/Q) = eQ$, and by Theorem 11, $D(G) \subseteq Q$ as asserted.

Definition. Let G be an infinite group, and let $\aleph_0 \leqq A \leqq o(G)$. Let $K(A, G)$ be the set of $k \varepsilon G$ such that $o(E(k)) < A$. Let $D(A, G)$ be the intersection of all subgroups G_α of G such that $o(G_\alpha) \geqq A$.

Note that $K(G) = K(o(G), G)$, $D(G) = D(o(G), G)$. If $A < B$, then $K(A, G)$ is a subgroup of $K(B, G)$ and $D(A, G)$ is a subgroup of $D(B, G)$. Moreover $K(A, G) \subseteq D(A, G)$, and $K(A, G)$ is fully characteristic and $D(A, G)$ is strictly characteristic in G. It follows from Corollary 2 to Theorem 8 that if $o(G) > A \geqq \aleph_0$, then $K(A, G)$ is a cyclic group of order p^n.

UNIVERSITY OF KANSAS.

REFERENCES.

[1] R. Baer, "The subgroup of the elements of finite order of an Abelian group," *Annals of Mathematics*, vol. 37 (1936), pp. 766-781.

[2] S. N. Černikov, "Infinite groups with finite layers," *Matematičeskiĭ Sbornik*, vol. 22 (1948), pp. 101-133 (Russian).

[3] L. Szélpál, "Die unendlichen Abelschen Gruppen mit lauter endlichen echten Untergruppen," *Publicationes Mathematicae (Debrecen)*, vol. 1 (1949), pp. 63-64.

ON ISOMETRIC SURFACES.*

By Aurel Wintner.

1. The starting point of the following considerations was the observation that, in the differential geometry of surfaces, the fundamental notion of *isometry* is frequently used in a loose and misleading sense. Even when writers are careful enough to specify the assumptions of smoothness (the actual degree of differentiability, possibly the analyticity) of the two surfaces concerned, usually no mention is made of the degree of smoothness required of the mapping of the two surfaces which realizes their isometry in question.

In order to clarify this objection, a few definitions will be needed. First, if S denotes a sufficiently small (open) piece of a surface in the X-space, where $X = (x, y, z)$, and, if on a sufficiently small (u^1, u^2)-domain of a parameter plane,

$$S: \quad X = X(u^1, u^2) \tag{1}$$

is a parametrization in which the vector function $X(u^1, u^2)$ is of class C^1 and such that the vector product of $X_1 = \partial X/\partial u^1$ and $X_2 = \partial X/\partial u^2$ does not vanish, then (1) is called a C^1-parametrization of S. By a C^n-parametrization is meant[1] a C^1-parametrization in which the function $X(u^1, u^2)$ is of class C^n. If S, when given as a set of points in the X-space, has *some* C^n-parametrization, then S will be called of class C^n.

In Section 4, a corresponding manner of speaking will be used with reference to the surface S and to the function class

$$C^n(\lambda), \qquad (0 < \lambda < 1), \tag{2}$$

where λ denotes a (locally uniform) Hölder index for the n-th derivatives of $X(u^1, u^2)$, for a fixed n. Similarly, (1) will be called an analytic parametrization if it is a C^1-parametrization having the property that the function $X(u^1, u^2)$ can be represented (locally) as a convergent power series in (u^1, u^2), while S will be called analytic if it has *some* analytic parametrization. Note that every analytic S has C^n-parametrizations which are not C^{n+1}-parametrizations, where n can be chosen arbitrarily.

* Received January 16, 1951; revised October 15, 1951.

[1] If $n = 1, 2, \cdots$. The limiting case $n = 0$, which (unless the contrary is implied), will be excluded, refers to one-to-one continuous parametrizations (1) of S.

Besides S, consider another S, say S'. Suppose that both S and S' are of class C^1, at least. If they are of class C^n, where $n=\infty$ is not excluded, let

$$(3) \qquad S:\ X(u^1,u^2) \quad \text{and} \quad S':\ X'(u'^1,u'^2)$$

be C^n-parametrizations; if S and S' are analytic, let their parametrizations (3) be chosen analytic. With the understanding that S and S' are sufficiently small, let

$$(4) \qquad u'^1=u'^1(u^1,u^2), \qquad u'^2=u'^2(u^1,u^2)$$

be a C^1-transformation (by this is meant that the two functions (4) are of class C^1 *and* have a non-vanishing Jacobian). If such a one-to-one transformation can be chosen so that it will transform the metric form, ds^2, on S into the metric form, ds'^2, on S', i. e., if the two vector functions (3) satisfy, for an appropriate choice of the mapping (4), the identity

$$(5) \qquad |\,dX'(u'^1,u'^2)\,| \equiv |\,dX(u^1,u^2)\,| \text{ by virtue of } (4),$$

then S and S' are called *isometric*. This is the definition which tacitly underlies the classical literature of the subject.

2. Needless to say, the C^1-character of the functions occurring in (3) and (4) makes (5) a meaningful statement, since *both* X and X' are functions of class C^1 in terms of (u^1,u^2) (or, equivalently, (u'^1,u'^2)) and possess therefore [2] the complete differentials [3] occurring in (5). On the other hand, even if the surfaces (3) are very smooth (say analytic), there is no justification for restricting (4) in (5) to transformations having a high degree of smoothness since nothing like such a restriction (say analyticity) is involved in the geometrical idea of a transformation (4) which preserves the metric, ds^2. This contrast leads, however, to geometrically undesirable situations.

In order to see this, consider the wording of the following assertion (stated, to be sure, because of its instructive nature only): "Two closed, convex, *analytic* surfaces, F and F', must be congruent [4] whenever they are (locally) isometric." What should be meant here by the assumption of the isometry of F and F'? The existence of

---

[2] This conclusion shows that, as far as (5) is concerned, the C^1-character of the admitted mappings (4) could slightly be generalized.

[3] Actually, not even this, the existence of the "vectors" dX, but only the existence of the "distances" $|\,dX\,|$, is needed for the formulation of the requirement (5) (so that isometry can be defined in a manner which is even more general than, but geometrically just as meaningful as, the definition referred to in the preceding footnote).

[4] Under the group consisting of the translations, rotations *and* reflections of the X-space.

(a) analytic transformations (4) or (b) just C^1-transformations (4)

which preserve the ds^2? If the statement is meant in its interpretation (a), then the content of the statement is hardly geometrical in nature, since it does not preclude the following possibility: The geometrical objects, F and F', need not be congruent if they are isometric under mappings (4) which are non-analytic but very smooth, say of class C^∞. Hence, in order to make the statement geometrically significant, its truth in interpretation (b) must be proved to imply its truth in interpretation (a).

It so happens that, in the theorem, quoted above, also the restriction of the convex surfaces F, F' is unnecessary, since Herglotz's result [4], as predicted by Weyl [7], has nothing to do with the analyticity, but only with a specific degree of differentiability (C^n), of the given pair of surfaces; cf. Section 4 below. But then the theorem has again two interpretations, one being the above (b) and another, say (a_n), an interpretation which would be an appropriate C^n-analogue of the above (a); cf. the end of Section 4 below.

3. There is however a classical instance of isometry in which, in contrast to the Weyl-Herglotz problem, the analyticity of the surfaces is precisely the issue. It is S. Bernstein's theorem, the statement of which runs as follows: "If S and an analytic S' of positive Gaussian curvature are isometric, then S is analytic." This is Bernstein's own formulation of his theorem (the italicized statement in [1], p. 434). But it is again not specified which of the two interpretations defined above, (a) or (b), is meant under "isometry," and so it is again necessary to point out that the weaker formulation, (b), implies the stronger one, (a); see Section 3 below.

In addition, the above wording of Bernstein's theorem fails to specify the degree of smoothness required of the given surface, S. In this regard, Bernstein's proof makes it clear that S is assumed to be of class C^3, rather than, as one might desire (and expect from the above wording), of class C^1 only. This comes about by Bernstein's use of his general theorem. according to which every function $z = z(x, y)$ satisfying a partial differential equation, of second order, of elliptic type, and having analytic coefficients, must be analytic whenever it is of class C^3. Accordingly, the improved version of the above wording of Bernstein's theorem is as follows:

If S is of class C^3 and of positive Gaussian curvature, and if there exists a C^1-transformation (4) *of S into an analytic S': $X'(u'^1, u'^2)$, satisfying* (5) *(where the parametrization S: $X(u^1, u^2)$ is of class C^3), then S is analytic.*

Bernstein's proof is based on an application of geodesic polar coordinates

([1], p. 435). But a perusal of his proof shows that the application of such particular (u^1, u^2)-parameters is unimportant. In fact, under the assumption that S has a C^1-parametrization (1), all that is needed is the existence of a C^1-transformation (4) satisfying (5) and having the property that, while the vector function $X(u^1, u^2)$ is supposed to be of class C^1 in (u^1, u^2), it becomes of class C^3 in (u'^1, u'^2) by virtue of (4). The rest then follows from Bernstein's general theorem on elliptic differential equations and from Section 4 below.

These remarks also show that the theorem can well be formulated so as to avoid the difficulties involved in the various concepts of an isometry; namely, as follows:

Let (g_{ik}) be a binary, symmetric matrix of analytic functions

$$g_{ik} = g_{ik}(u^1, u^2)$$

and suppose that the curvature [5] of

$$ds^2 = g_{ik}(u^1, u^2)\, du^i du^k \tag{6}$$

is positive. Then a vector function $(x, y, z) = X = X(u^1, u^2)$ satisfying

$$| dX(u^1, u^2)|^2 \equiv ds^2 \tag{7}$$

must be analytic whenever it is of class C^3.

While this theorem might be true even if the last C^3 is relaxed to C^1, its truth or falsehood is undecided even if the C^3 is relaxed just to C^2 (in which latter case, but not in the C^1-case, (1) must have a curvature not only by virtue of the Theorema Egregium but in terms of the normal image of S as well).

4. If S and S' are of class C^n [analytic], let all their parametrizations (3) considered be restricted to C^n-parametrizations [analytic parametrizations], and let (4) be called a C^m-isometry [an analytic isometry] of the pair (3) if (5) is satisfied and (4) is a C^m-transformation [an analytic transformation] (in the sense that both functions occurring in (4) are of class C^m [analytic] with a non-vanishing Jacobian).

It is understood that, in the non-analytic case of this definition, $1 \leqq n \leqq \infty$ and $1 \leqq m \leqq \infty$, and that $m \neq n$ is allowed. In the analytic case, the above critique (Sections 2-3) will be disposed of by the following lemma:

If S and S' are analytic and C^1-isometric, then all of their C^1-isometries are analytic isometries.

[5] Calculated from the Theorema Egregium.

This lemma is a corollary of (i) the fact that, according to Gauss [2], every analytic S admits of an analytic parametrization (1) in which the line-element (6) appears in the normal form

$$ds^2 = g(u^1, u^2)\{(du^1)^2 + (du^2)^2\}, \qquad (g = g_{ii} > 0, g_{12} = 0);$$

to be combined with (ii) the fact that if two real-valued functions (4) are of class C^1 and satisfy the Cauchy-Riemann equations, then $u'^1 + iu'^2$ is an analytic function of $u^1 + iu^2$.

First, since S and S' are analytic, two applications of (i) supply the existence of two analytic transformations (of non-vanishing Jacobian), say of

$$\alpha:\ (u^1, u^2) \to (v^1, v^2) \text{ and } \beta:\ (u'^1, u'^2) \to (v'^1, v'^2),$$

by virtue of which the respective line-elements $|\,dX(u, v)\,|^2$ and $|\,dX'(u', v')\,|^2$ on S and S' become of the form

$$ds^2 = h(v^1, v^2)\{(dv^1)^2 + (dv^2)^2\} \text{ and } ds'^2 = h'(v'^1, v'^2)\{(dv'^1)^2 + (dv'^2)^2\},$$

where $h > 0$, $h' > 0$. On the other hand, since S and S' are supposed to be C^1-isometric, there exist two functions (4), of class C^1 and of non-vanishing Jacobian, satisfying

$$ds^2 = ds'^2 \text{ by virtue of } \gamma:\ (u^1, u^2) \to (u'^1, u'^2).$$

But the last three formula lines show that $\beta^{-1}\gamma\alpha^{-1}$ is a conformal transformation of a domain in the Euclidean (v^1, v^2)-plane into a domain in the Euclidean (v'^1, v'^2)-plane, and $\beta^{-1}\gamma\alpha^{-1}$ is a C^1-transformation, since γ is. It follows therefore from (ii) (where every u must be replaced by the corresponding v) that $\beta^{-1}\gamma\alpha^{-1}$ is analytic. In view of the analyticity of β and α, this proves that γ is analytic, which is the assertion of the above lemma.

One might think that this proof also leads to a C^n-analogue of the above lemma, with "conformal" transformations α and β which, instead of being analytic, as above, are of class C^n. Actually, this approach fails, since it applies only to the classes (2) (with an unspecified λ). In fact, Lichtenstein's analogue [6] of the above (i) is as follows: If $n > 0$ and $0 < \lambda^* < \lambda < 1$, then every S of class $C^n(\lambda)$ can conformally be mapped on a domain in the Euclidean plane by a transformation (4) of class $C^n(\lambda^*)$. But it is not known (and it is probably not true; cf. Section 5 below) that this remains true if both $C^n(\lambda)$ and $C^n(\lambda^*)$ are replaced by C^n itself.[7] Nevertheless, the straight

[6] Cf. [5], where the proof is given for $n = 1$ only but, as easily realized, applies for $n > 1$ also.

[7] Or, for that matter, if $0 < \lambda^* < \lambda < 1$ is replaced by $0 < \lambda^* = \lambda < 1$.

C^n-analogue of the above analytic lemma might be true, since all that follows is the failure of *conformal* normal forms.

5. The question raised at the end of Section 4, concerning the necessity of a Hölder index λ, can slightly be generalized, by omitting the restriction that the binary Riemannian line-element (6) be "embedded" in the X-space, as required by (1) and (7). Then, if (u^1, u^2), (v^1, v^2) are denoted by (u, v), (p, q), and g_{11}, g_{12}, g_{22} by E, F, G, respectively, the question becomes, for every fixed positive n (including $n-1=0$) the following:

$(?_n)$ If E, F, G are given functions of class C^{n-1} on a sufficiently small (u, v)-domain, and if

$$EG - F^2 > 0 \tag{8}$$

(that is, if the quadratic form

$$ds^2 = E(u, v)\,du^2 + 2F(u, v)\,du\,dv + G(u, v)\,dv^2 \tag{9}$$

is definite), then must there exist two functions

$$p = p(u, v), \qquad q = q(u, v), \tag{10}$$

of class C^n and of non-vanishing Jacobian, with the property that the form (9) becomes

$$ds^2 = D(p, q)\,(dp^2 + dq^2) \tag{11}$$

by virtue of (10) (for some function $D > 0$, determined by the Jacobian of (10))?

The answer to $(?_1)$ is affirmative if and only if (8) and the mere continuity $(C^{n-1} = C^0)$ of the three coefficient functions of (9) are always sufficient to assure the existence of at least one C^1-solution (10) (of non-vanishing determinant) for Beltrami's form of the Cauchy-Riemann equations, that is, for

$$q_u = (Fp_u - Ep_v)/W, \qquad q_v = -(Fp_v - Gp_u)/W, \tag{12}$$
$$\text{where } W^2 = EG - F^2.$$

If a solution (10) of (12), instead of being of class C^1 only, is of class C^2, then, since the rule $\phi_{uv} = \phi_{vu}$ becomes applicable to both functions (10), it follows from (12), and from the non-vanishing of the Jacobian of (10), that both $\phi = p$ and $\phi = q$ are solutions $\phi = \phi(u, v) \neq \text{const.}$ of

$$\{(G\phi_u - F\phi_v)/(EG - F^2)^{\frac{1}{2}}\}_u + \{(E\phi_v - F\phi_u)/(EG - F^2)^{\frac{1}{2}}\}_v = 0 \tag{13}$$

(Laplace-Beltrami). Hence, if the answer to the case $n=2$ of the question $(?_n)$ is affirmative, then the (homogeneous, linear, elliptic) partial differential equation (13) must possess some non-constant solution $\phi=\phi(u,v)$ of class C^2 whenever $E(u,v)$, $F(u,v)$, $G(u,v)$ are functions of class C^1 satisfying (8) (incidentally, this can be concluded without a detour through (12) also; cf. [6], pp. 1295-1297). But it is unlikely that this (hence an affirmative answer to the question $(?_n)$ if $n=2$) should turn out to be true. The reason for being skeptical is as follows:

Choose

$$(14) \qquad E(u,v)\equiv 1 \text{ and } F(u,v)\equiv 0$$

(so that u, v are, in the main, "geodesic polar coordinates" in the sense of Gauss, with reference to the metric (9) and to a point of the (u,v)-plane). It is clear from (14) that (8) is satisfied if $G>0$, that (12) reduces to

$$(15) \qquad p_v=-g(u,v)q_u, \qquad q_v=g(u,v)p_u,$$

if g denotes $G^{\frac{1}{2}}>0$, and that (13) therefore simplifies to

$$(16) \qquad (\phi_u g)_u+(\phi_v/g)_v=0.$$

Hence, if the answer to the question $(?_2)$ is affirmative, then (16) must have a solution $\phi=\phi(u,v)\neq$ const. of class C^2 whenever $g=g(u,v)$ is a positive function of class C^1. But then the coefficients of (16), being composed of g_u, g_v and g, are just continuous, and so (16) does not seem to be substantially different from the differential equation

$$(17) \qquad \phi_{uu}+\phi_{vv}+f(u,v)\phi=0,$$

in which the given function, f, is just continuous. However, it was shown in [8] that it is possible to choose a continuous function $f(u,v)>0$ in such a way that (17) will fail to have *any* (continuous) solution $\phi(u,v)\neq$ const. on *any* (u,v)-domain.

6. Since the answer to the questions $(?_n)$ is not known, it is worth mentioning that the answer is surely in the negative if (8) is replaced by

$$(18) \qquad EG-F^2<0$$

and, correspondingly, (11) by

$$(19) \qquad ds^2=D(p,q)(dp^2-dq^2).$$

For the case $n=1$ of the respective class C^{n-1}, C^n (a case in which E, F, G

are just continuous, hence (10) is just a C^1-transformation), this can, in other words, be formulated as follows: There exist *hyperbolic line-elements* (9), *with continuous coefficient functions* E, F, G, *which cannot be mapped "conformally" on the non-Euclidean* (p, q)-*plane* by any mapping (10) of class C^1.

In order to prove this, choose (9) as in (14). Then (18) reduces to $G < 0$ and, since (11) is replaced by (19), what corresponds to the Cauchy-Riemann system (15) becomes

$$(20) \qquad p_v = f(u, v) q_u, \qquad q_v = f(u, v) p_u,$$

where $f = (-G)^{\frac{1}{2}} > 0$. Hence the assertion is that, if $f(u, v)$ is an arbitrary positive continuous function, then (20) need not have any solution (10) of class C^1, unless the Jacobian $\partial(p, q)/\partial(u, v)$ vanishes identically.

In order to prove the existence of such an f, write (20) in the form $(p \pm q)_v = \pm f(u, v)(p \pm q)_u$. The latter shows that $\phi = p + q$ must satisfy

$$(21) \qquad \phi_v - f(u, v)\phi_u = 0$$

(and $\phi = p - q$ the homogeneous, linear, partial differential equation which results from (21) if f is replaced by $-f$). But it was shown in [8], pp. 733-734. that there exist continuous functions $f = f(u, v) > 0$ having the property that, no matter where, and no matter how small, a (u, v)-circle be chosen, the differential (21) will not possess within the circle any solution $\phi = \phi(u, v)$ of class C^1, except the trivial solution $\phi(u, v) = \text{const.}$ In view of the connection between (21) and (20), this proves more than what was claimed above for the case $n - 1 = 0$.

7. The existence of an f, satisfying the conditions used in connection with (20), has a significance from the point of view of the theory of multipliers (Euler). In order to see this, let P and Q be continuous on a simply connected open (x, y)-domain, and let J denote any smooth Jordan curve, finally D any open set contained in this domain. If $\mu(x, y)$ is a continuous function on the latter, and if the line integral, along every C, of $\mu(x, y)$ times

$$(22) \qquad P(x, y)\,dx + Q(x, y)\,dy$$

vanishes, then $\mu(x, y)$ will be called a multiplier of the Pfaffian (22), provided that $\mu(x, y)$ does not vanish identically. It turns out that, if no assumptions of smoothness (involving partial derivatives) are placed on the functions $P(x, y), Q(x, y)$, then *it is possible to choose the Pfaffian* (22), *with con-*

tinuous $P(x,y)$ *and* $Q(x,y)$, *in such a way that on no open* (x,y)-*domain* D *will* (22) *possess a multiplier.*

In order to see this, suppose that (22) is such as to possess a multiplier, $\mu(x,y)$. Then the line integral of $\mu(x,y)$ times the Pfaffian (22), when extended from a fixed (x_0,y_0) to a variable (x,y) along a smooth Jordan arc, will be a function of (x,y) and, if this function is denoted by $\phi(x,y)$, the partial derivatives ϕ_x, ϕ_y will exist and satisfy the relations

$$\phi_x(x,y)=\mu(x,y)P(x,y),\qquad \phi_y(x,y)=\mu(x,y)Q(x,y). \tag{23}$$

In addition, ϕ_x and ϕ_y are continuous. This follows from (23), since P, Q and μ are supposed to be continuous. Accordingly, $\phi(x,y)$ is of class C^1.

Choose $P(x,y)=1$ for every (x,y), put $Q=f$, and write u and v instead of y and x, respectively. Then (23) becomes

$$\phi_v(u,v)=\mu(u,v),\qquad \phi_u(u,v)=\mu(u,v)f(u,v) \tag{24}$$

(if u and v are interchanged as arguments). Substitution of $\mu(u,v)$ from the first of the relations (24) into the second shows that $\phi(u,v)$ is a solution of (21). Hence, if $f(u,v)$ is so chosen as at the end of Section 6, then, since $\phi(u,v)$ is of class C^1, it follows that $\phi(u,v)=$ const. on every D.

In particular, the partial derivative $\phi_v(u,v)$ vanishes identically. In view of the first of the relations (24), this means that μ vanishes identically. Since such a μ was excluded in the definition of a multiplier, the proof is complete.

8. In Section 2, reference was made to Herglotz's theorem [4], which states that, under certain assumptions of smoothness, two closed, essentially convex [8] surfaces are congruent whenever they are isometric ("congruence" is meant in the sense including "anti-congruence," i. e., reflections on a plane of the Euclidean X-space are allowed). With regard to the assumptions of smoothness, the situation is as follows: While the so-called derivation formulae (those of Gauss and Weingarten) hold on any surface of class C^2, Herglotz's proof involves differentiations (of first order) of these formulae and assumes therefore that the convex surfaces are of class C^3. In addition, the proof depends on a tacit assumption, one corresponding to the comments in Sections 1-2 above. In fact, the differentiated formulae, just mentioned, contain local

[8] A convex surface (of class C^2) will be called essentially convex if the set of its parabolic points (if there are any) is of Lebesgue measure 0; cf. (52) below. For instance, convexity implies essential convexity under the proviso that the Gaussian curvature should be positive with the possible exception of isolated points or curves.

representations of the two C^3-surfaces in terms of the *same* parameter plane (u^1, u^2). According to the terminology introduced at the beginning of Section 4, this additional hypothesis means that the two surfaces are supposed to be locally C^3-isometric.

From a geometrical point of view, there is an objection to most C^3-assertions in the theory of surfaces. In fact, the curvatures (total and mean, K and H) and even both fundamental forms ($g_{ik}du^i du^k$ and $h_{ik}du^i du^k$) exist, and are continuous, on surfaces of class C^2. Correspondingly, the restriction of a theorem to surfaces of class C^3, instead of to more inclusive (and *geometrical*) class C^2, is often due to accidental formal difficulties, resulting from the limitations of the underlying analytical tools, rather than to the actual geometrical situation.

For instance, while the fundamental existence theorem of the differential geometry of surfaces (Bonnet) is a C^3-theorem in its classical wording, it can, with some effort, be freed (cf. [3], pp. 758-760) of the unnatural C^3-restriction. In what follows, the possibility of a corresponding reduction, $C^3 \rightarrow C^2$, will be proved in Herglotz's theorem, both with regard to the smoothness of the two convex surfaces and that of their underlying isometry. In other words, it will be proved that the content of the theorem can be refined to the following statement:

Two closed, essentially[9] *convex surfaces of class C^2 are congruent whenever they are (locally) C^2-isometric.*

It would be desirable to reduce the theorem even further, by showing that every (local) C^1-isometry of two C^2-surfaces is a C^2-isometry by necessity (cf. Section 4).

9. As in (3), let S and S' be two, sufficiently small, pieces of surfaces both of which are of class C^2. While in the application of the lemma to be derived, S and S' represent pieces of two closed, convex surfaces, no such additional assumption is made now (so that the Gaussian curvatures need not be non-negative). Suppose that S and S' are C^2-isometric. Then after a suitable C^2-transformation (4) (of non-vanishing Jacobian), (3) can be assumed to be in the form

$$S\colon X(u^1, u^2) \text{ and } S'\colon X'(u^1, u^2), \tag{25}$$

where

$$| dX(u^1, u^2)|^2 = | dX'(u^1, u^2)|^2 \tag{26}$$

[9] See the preceding footnote.

(since (4) in (5) is now $u'^1 = u^1$, $u'^2 = u^2$), and where the vector functions (25) are of class C^2 and such that the vector products $[X_1, X_2]$, $[X'_1, X'_2]$ of the respective partial derivatives $Z_i = \partial Z/\partial u^i$ $(i = 1, 2; Z = X, X')$ do not vanish. Thus there exist normal unit vectors, say

$$(27) \qquad N = [X_1, X_2]/|[X_1, X_2]| \text{ and } N' = [X'_1, X'_2]/|[X'_1, X'_2]|,$$

and, according to (26), the scalar products

$$(28) \qquad g_{ik} = X_i \cdot X_k, \qquad g_{ik} = X'_i \cdot X'_k$$

are respectively identical (and have a determinant which is positive, say

$$(29) \qquad g > 0, \text{ where } g = (\det g_{ik})^{\frac{1}{2}},$$

since $[X_1, X_2] \neq 0$), and the functions (27), (28), (29) of (u^1, u^2) are of class C^1. In contrast to (28), expressing the identity of the two first fundamental forms, $g_{\alpha\beta}du^\alpha du^\beta$ and $g'_{\alpha\beta}du^\alpha du^\beta$, there are two distinct second fundamental forms, the coefficients of which are defined by

$$(30) \qquad h_{ik} = N \cdot X_{ik}, \qquad h'_{ik} = N' \cdot X'_{ik}$$

$(X_{ik} = \partial^2 X/\partial u^i \partial u^k)$, and, since the functions (25) and (27) are of class C^2 and C^1, respectively, the functions (30) are just continuous.

The two matrices (h_{ik}), (h'_{ik}) (which can be definite, semi-definite or indefinite) have the same determinant at every point (u^1, u^2); in other words,

$$(31) \qquad g^2 K = \det h_{ik}, \qquad g^2 K = \det h'_{ik},$$

if (29) and the first of the relations (31) are considered as the definition of the continuous function $K = K(u^1, u^2) \gtreqless 0$. If the function (25) were of class C^3, hence the functions (28) of class C^2, then, since (31) defines the Gaussian curvatures, K, of S and S', respectively, the identity claimed by (31) would follow from the classical form of the Theorema Egregium (note that the latter contains the second derivatives of functions g_{ik}). This proof of (31) fails to apply, since the functions g_{ik} are supposed to be of class C^1 only. That (31) is nevertheless true in the present case also, follows from the circumstance that, due to a fact first observed by Weyl, a certain integrated form of the Theorema Egregium happens to hold for every surface of class C^2 (for references and for a simple proof, cf. [3], p. 760 and formula (7) on p. 759).

Besides the (common) Gaussian curvature, K, of S and S', consider their (generally different) mean curvatures,

$$(32) \qquad H = \tfrac{1}{2} g^{\alpha\beta} h_{\alpha\beta}. \qquad\qquad (33) \qquad H' = \tfrac{1}{2} g^{\alpha\beta} h'_{\alpha\beta}.$$

where (g^{ik}) denotes the reciprocal matrix, $(g_{ik})^{-1}$, of (g_{ik}). Put

$$(34) \qquad 2g^2J = h_{11}h'_{22} - 2h_{12}h'_{12} + h_{22}h'_{11}.$$

In view of (29), the function $J = J(u^1, u^2)$ defined by (34) is the "mixed" form (in the sense of Brunn and Minkowski) of the two expressions (31). (This "mixed" Gaussian curvature seems first to have arisen in connection with Weingarten's "associated surfaces" of his theory of infinitesimal deformations; cf. the reference in footnote [10] below.) All three functions (32), (33), (34) are continuous, since the functions (30), (28) are.

For the above-defined functions, the integral relation

$$(35) \qquad \int_C g^{-1}(h'_{2a}X \cdot X_1 - h'_{1a}X \cdot X_2)du^a = 2\int\int_D g\{(N \cdot X)J + H'\}du^1du^2$$

is an identity in C, where C on the left denotes a piecewise smooth, Jordan curve contained in the (sufficiently small and hence, without loss of generality, simply connected) parametric $(u^1. u^2)$-domain on which the surfaces (25) are given, while D on the right of (35) denotes the interior of C. This is Herglotz's fundamental identity, proved by him under his C^3-assumption for (25). It will be shown that (35) holds under the present C^2-assumption also. The C^2-theorem on closed, convex surfaces, as announced in Section 8, will then follow from the "local" identity (35), since the balance of Herglotz's proof remains unaltered.

10. The derivation formulae of Gauss and Weingarten,

$$(36) \qquad X_{ik} = \Gamma^a_{ik}X_a + h_{ik}N$$

and

$$(37) \qquad N_i = - g^{a\beta}h_{ia}X_\beta$$

(cf. [3], p. 758), where the $\Gamma^i_{jk} = \Gamma^i_{jk}(u^1, u^2)$ are Christoffel's symbols, hold on every surface of class C^2. On the other hand, the Mainardi-Codazzi equations (representing that part of the integrability conditions of the system (36)-(37) which remains after a satisfaction of the Theorema Egregium) cannot be applied on a surface of class C^2, since they contain the derivatives of the functions h_{ik}, whereas the latter functions are of class C^1 only if the surface is of class C^3. It was however shown in [3], pp. 759-760, that if the surface is of class C^2, then the Mainardi-Codazzi equations still apply in their "integrated" form,

$$(38) \qquad \int_C h_{ia}du^a = \int\int_D (\Gamma^a_{i1}h_{a2} - \Gamma^a_{i2}h_{a1})\,du^1du^2, \qquad (i = 1, 2),$$

where the ("arbitrary") Jordan curve C and its interior D are restricted only by the assumptions which were specified for (35) above.

Let (36′), (37′), (38′) denote the relations which result from (36), (37), (38) if X, N, h_{ik} are replaced by X', N', h'_{ik}, respectively (note that $\Gamma' = \Gamma$, since $g'_{ik} = g_{ik}$).

Since the function (29) is of class C^1, an application of the general Lemma of [3], p. 761, shows that (38) implies the relations

$$(39) \qquad \int_C g^{-1} h'_{i\alpha} du^\alpha = \int\int_D g^{-1} \Delta_i du^1 du^2, \qquad (i = 1, 2),$$

if the (continuous) functions Δ_1, Δ_2 are defined by

$$(40) \qquad \Delta_i = (-1)^j (\Gamma^j_{11} h'_{22} - 2\Gamma^j_{12} h'_{12} + \Gamma^j_{22} h'_{11}), \qquad j \neq i$$

(so that $j = 2$ or $j = 1$ according as $i = 1$ or $i = 2$). The same general lemma also shows that the partial derivative, $x_2 = \partial x / \partial u^2$, of the first component of the vector $X = (x, y, z)$ can be "inserted" into the case $i = 1$ of the relation (39), and that this leads to

$$\int_C g^{-1} x_2 h'_{1\alpha} du^\alpha = \int\int_D g^{-1} (\Delta_1 x_2 + h'_{12} x_{12} - h'_{11} x_{22}) du^1 du^2.$$

Similarly,

$$\int_C g^{-1} x_1 h'_{2\alpha} du^\alpha = \int\int_D g^{-1} (\Delta_2 x_1 + h'_{22} x_{11} - h'_{12} x_{12}) du^1 du^2.$$

Hence, by subtraction,

$$(41) \qquad \int_C g^{-1} (x_1 h'_{2\alpha} du^\alpha - x_2 h'_{1\alpha} du^\alpha) = \int\int_D [\cdot\ \cdot\ \cdot] du^1 du^2,$$

where the expression on the right is the difference of the double integrals occurring in the preceding two formulae.

Corresponding to $X = (x, y, z)$, let $N = (a, b, c)$. Then, according to (36),

$$x_{ik} = \Gamma^\alpha_{ik} x_\alpha + a h_{ik}.$$

Hence, if x_{11}, x_{12}, x_{22} are multiplied by $g^{-1} h'_{22}$, $-2g^{-1} h'_{12}$, $g^{-1} h'_{11}$, respectively, it is seen, by addition, that the difference which on the right of (41) was indicated by $[\cdot\ \cdot\ \cdot]$ can be written as

$$[\cdot\ \cdot\ \cdot] = g^{-1} a (h_{11} h'_{22} - 2h_{12} h'_{12} + h_{22} h'_{11}),$$

since the functions Δ_1, Δ_2 occurring in the definition of the difference $[\cdot\cdot\cdot]$ are given by (40).

According to (34), the preceding representation of $[\cdot\cdot\cdot]$ can be written as $[\cdot\cdot\cdot] = 2gaJ$. Thus (41) becomes

$$(42) \qquad \int_C g^{-1}(x_1h'_{2a}du^a - x_2h'_{1a}du^a) = \int\int_D 2gaJ\,du^1du^2,$$

where a denotes the direction cosine defined by $N = (a, b, c)$. If the general Lemma of [3], p. 761, is applied again, this time in order to "insert" x as a factor gaJ on the right of (42), it follows that (42) implies the relation

$$\int_C g^{-1}x(x_1h'_{2a}du^a - x_2h'_{1a}du^a)$$
$$= \int\int_D \{2gaxJ + g^{-1}[x_1(x_1h'_{22} - x_2h'_{12}) - x_2(x_1h'_{12} - x_2h'_{11})]\}du^1du^2.$$

But x, a can here be replaced by any of the respective components of $X = (x, y, z)$, $N = (a, b, c)$. Hence the last relation, when compared with the definitions (28) and (32), completes the proof of (35).

11. It was mentioned at the end of Section 9 that it is sufficient to prove (35) under the C^2-assumption in (25). Actually, since the end of Herglotz's argument depends on an appeal to the fundamental uniqueness theorem of local differential geometry, and since the classical formulation of this theorem (Bonnet) is confined to surfaces of class C^3, for the sake of completeness the balance of the proof will also be given. (Incidentally, it will be worth noting that, just as the Gauss-Bonnet representation of the genus, the classical formula for the surface average of the mean curvature and Herglotz's generalization of this formula for the "mixed" case, namely (43) and (44) below, are not restricted to surfaces of genus 0; cf. [4], p. 128.)

Let F and F' be two orientable, closed, homeomorphic, locally C^2-isometric surfaces of class C^2. Draw on F an oriented net, and on F' the corresponding net, of piecewise smooth Jordan curves in such a way that, if $S_1, \cdots, S_h$ and $S', \cdots, S'_h$ denote the interiors of these (oriented) Jordan curves, then, on the one hand, every pair $(S, S') = (S_j, S'_j)$, where $j = 1, \cdots, h$, has a C^2-representation of the form (25) and, on the other hand, (35) holds for every $C = C_j$ and for the corresponding $D = D_j$, where D_j denotes the (u^1, u^2)-domain to which S_j and S'_j are referred in (25). Then, if (35) is applied to every j, summation with respect to j gives

$$0 = \int\int_E g\{(N \cdot X)J + H'\} du^1 du^2,$$

where $E = D_1 + \cdots + D_k$. This step assumes that the line integrals, which are cancelled by the addition, have a geometrical meaning (in the sense that they are independent of the different parametrizations, used on the different pieces S_j), which, however, can easily be ascertained.

Let dF and dF' denote the surface elements on F and F', respectively, and put

$$p = N \cdot X. \tag{43}$$

Then the preceding integral relation can be written in the form

$$\int\int_F JpdF = -\int\int_{F'} H'dF', \tag{44}$$

since, according to (29) and (28), both dF and dF' are identical with gdu^1du^2. It follows that

$$\int\int_F LpdF = \int\int_{F'} H'dF' - \int\int_F HdF, \tag{45}$$

if L is defined by

$$2g^2L = \det(h'_{ik} - h_{ik}) \qquad (g^2 = \det g_{ik} > 0). \tag{46}$$

In fact, it is seen from (31)-(33) and (46) that

$$K - J = L. \tag{47}$$

But if F' is particularized to F, then (46) reduces to $L = 0$, hence (47) to $K = J$, and therefore (44) to

$$\int\int_F KpdF = -\int\int_F HdF. \tag{48}$$

Finally, (45) follows by subtracting (44) from (48) and using (47).

12. The closed, orientable surfaces F, F' have thus far been of arbitrary genus. It will be supposed that the Gaussian curvature $K = K(u^1, u^2)$ is positive throughout or, more generally, that

$$K > 0 \text{ almost everywhere} \tag{49}$$

(hence $K \geqq 0$ everywhere) on F and/or F'. Then, according to Hadamard,

the genus must be 0, and so F, F' are closed convex surfaces. In particular, since (43) is the function of support for F, it can be assumed, (by choosing the origin of the X-space at a point which is not on F) that $p = p(u^1, u^2)$ is positive at every point of F.

A repetition of the argument used by Herglotz now shows that

$$h'_{ik} = h_{ik} \tag{50}$$

holds as an identity. His proof of this identity can be modified as follows:[10] It is readily verified that if $a_{\alpha\beta}x^\alpha x^\beta$, $b_{\alpha\beta}x^\alpha x^\beta$ are two positive definite, binary, quadratic forms of common determinant, $\det a_{ik} = \det b_{ik} > 0$, then their difference is either an indefinite or the null form (i. e., either $\det(a_{ik} - b_{ik}) < 0$ or $(a_{ik}) = (b_{ik})$ must hold).

In view of (49) and (31), the assumptions of this alternative are satisfied by $a_{ik} = h_{ik}(u^1, u^2)$, $b_{ik} = h'_{ik}(u^1, u^2)$ at almost every point (u^1, u^2). It follows therefore from (49) that, if a (u^1, u^2)-set of measure 0 is disregarded, then either $L < 0$ or (50) holds at each of the remaining points (in particular, $L \leqq 0$ holds everywhere). Hence it is seen from (49) and (45), where $p > 0$, that

$$0 \geqq \iint_{F} H'dF' - \iint_{F} HdF. \tag{51}$$

For reasons of symmetry, (51) must remain true if F and F' are interchanged. Consequently, the last inequality must actually be an equality. It follows therefore from (45) that

$$\iint_{F} LpdF = 0. \tag{52}$$

Since $L \leqq 0$ and $p > 0$ hold everywhere, (52) implies that $L = 0$ holds almost everywhere, and therefore, by continuity, everywhere. In view of (46), this proves (50).

In order to complete the proof of the theorem, it is only necessary to apply to (28) and (50) the local uniqueness theorem of [3] (Theorem (I), p. 760), which states that the first and second fundamental forms of a surface S *of class* C^2 determine S uniquely.

THE JOHNS HOPKINS UNIVERSITY.

---

[10] This variant of the corresponding explicit calculation in Herglotz's paper seems to be well-known. It was used by Professor Heinz Hopf in a lecture given this spring at Johns Hopkins, and it was known to the writer since he first read Herglotz's paper. Cf. also L. Bianchi, *Vorlesungen über Differentialgeometrie* (ed. 1899), the last quarter of p. 293 and formula (8*).

REFERENCES.

[1] S. Bernstein, "Sur la déformation des surfaces," *Mathematische Annalen*, vol. 60 (1905), pp. 434-436.

[2] C. F. Gauss, *Werke*, vol. 4 (1873), pp. 193-216.

[3] P. Hartman and A. Wintner, "On the fundamental equations of differential geometry," *American Journal of Mathematics*, vol. 72 (1950), pp. 757-774.

[4] G. Herglotz, "Ueber die Starrheit der Eiflächen," *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, vol. 15 (1943), pp. 127-129.

[5] L. Lichtenstein, "Zur Theorie der konformen Abbildung. Konforme Abbildung nichtanalytischer, singularitätenfreier Flächenstücke auf ebene Gebiete," *Bulletin International de l'Académie des Sciences de Cracovie, Classe des Sciences Mathématiques et Naturelles*, ser. A, 1916, pp. 192-217.

[6] ———, "Neuere Entwicklung der Theorie partieller Differentialgleichungen zweiter Ordnung vom elliptischen Typus," *Encyklopädie der Mathematischen Wissenschaften*, II C 12 (1924), pp. 1277-1334.

[7] H. Weyl, "Ueber die Bestimmung einer geschlossenen konvexen Fläche durch ihr Linienelement," *Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich*, vol. 61 (1915), pp. 40-72.

[8] A. Wintner, "On the Hölder restrictions in the theory of partial differential equations," *American Journal of Mathematics*, vol. 72 (1950), pp. 731-738.

ON UNSMOOTH TWO-DIMENSIONAL RIEMANNIAN METRICS.*

By Philip Hartman.**

This paper will be concerned with two quite different questions. Part I will deal with the question of the embedding into 3-dimensional Euclidean space a 2-dimensional Riemannian metric which has the Tchebychef normal form $du^2 + 2\cos\phi\, dudv + dv^2$, where it is assumed that this form has the "curvature" -1, but it is only assumed that ϕ is continuous. It will be shown that such an embedding exists and is unique, up to Euclidean movements, when $u =$ const. and $v =$ Const. are required to be asymptotic lines.

In Part II, it will be shown that if a positive-definite form $g_{ik}du^i du^k$, where $i, k = 1, 2$, of class C' is transformed into another form $G_{ik}dU^i dU^k$ with the same properties by a transformation $u^i = u^i(U^1, U^2)$ of class C', then the transformation is necessarily of class C''. This result is used to establish the uniqueness statement of Part I and has other applications, cf. § 7 below.

Part I.

1. Let $z = z(x, y)$, defined in a vicinity of $(x, y) = (0, 0)$, be a surface S of class C'' and possess the Gaussian curvature $K = -1$. It has been shown [4] that these conditions on S are sufficient to assure the existence of a transformation, defined in a vicinity of $(u, v) = (0, 0)$,

$$(1) \qquad x = x(u, v), \qquad y = y(u, v), \qquad (x(0,0) = y(0,0) = 0),$$

of class C', with non-vanishing Jacobian and with the properties that, in the resulting parametrization $X = X(u, v) = (x(u, v), y(u, v), z(x(u, v), y(u, v)))$ of S, the arcs $u =$ const. and $v =$ Const. are asymptotic curves and the squared element of arc-length on S has the Tchebychef form

$$(2) \quad ds^2 = g_{ik}(u, v)\, du^i du^k = du^2 + 2\cos\phi\, dudv + dv^2, \quad (u^1, u^2) = (u, v).$$

Since $\det(g_{ik}) = \sin^2\phi$, it follows that

$$(3) \qquad \sin\phi \neq 0, \qquad (\phi = \phi(u, v)).$$

* Received February 1, 1951.

** John Simon Guggenheim Memorial Foundation Fellow, on leave of absence from The Johns Hopkins University.

Finally, the function ϕ satisfies the Hazzidakis relation

$$(4)\qquad [\phi]=\int_{u_1}^{u_2}\int_{v_1}^{v_2}\phi(u,v)\,du\,dv,\ \text{ where }\ u_1\leqq u\leqq u_2,\ v_1\leqq v\leqq v_2$$

is any rectangle on which (1) is defined and

$$(5)\qquad [\phi]=\phi(u_1,v_1)-\phi(u_2,v_1)+\phi(u_2,v_2)-\phi(u_1,v_2).$$

Since (1) is of class C', the function ϕ is continuous. It was also shown in [4] that ϕ is of class C' if and only if $z=z(x,y)$ is of class C'''. (The existence of surfaces $S: z=z(x,y)$ which are of class C'', without being of class C''', and which are pseudo-spheres ($K=-1$) will be clear from the considerations below.)

In this paper, a converse of the above result will be considered.

(*) *Let $\phi=\phi(u,v)$ be a continuous function on the rectangle*

$$(6)\qquad R: |u|\leqq a,\quad |v|\leqq b,$$

satisfying the inequality $0<\phi<\pi$ and the relation (4) *for all rectangles in* (6). *Then, for sufficiently small x, y, there exists one and, up to Euclidean movements of the (x,y,z)-space, only one pseudo-sphere $S: z=z(x,y)$ of class C'' which belongs to $\phi(u,v)$ in the sense of the paragraph above.*

2. In order to make clear the content of this assertion, suppose first that ϕ is of class C', then (4) is equivalent to the existence and continuity of the second mixed partial derivative $\phi_{uv}=\phi_{vu}$ and to

$$(7)\qquad \phi_{uv}=\sin\phi.$$

Suppose further that ϕ is even smoother, say of class C'', then one expects the corresponding $X=X(u,v)$ to be of class C'''. Since $u=\text{const.}$ and $v=\text{Const.}$ are to become asymptotic arcs, the diagonal elements $h_{11}(u,v)$, $h_{22}(u,v)$ of the second fundamental matrix will be zero. Also, $K\equiv-1$ means $\det(h_{ik})\equiv-\det(g_{ik})=-\sin^2\phi$; so that $h_{12}{}^2=\sin^2\phi$. Define $h_{ik}=h_{ik}(u,v)$ by

$$(8)\qquad h_{ik}du^i du^k=2\sin\phi\,du\,dv.$$

The other choice, $h_{ik}du^idu^k=-2\sin\phi\,du\,dv$, merely corresponds to the enumeration $(u^1,u^2)=(v,u)$ (rather than to $(u^1,u^2)=(u,v)$).

It is easily verified that (g_{ik}) and (h_{ik}) given by (2) and (8), respectively, which are of class C'', satisfy the integrability conditions of Gauss and

Mainardi-Codazzi by virtue of (7). Thus, the standard theorem of Bonnet supplies a surface $S: X = X(u, v)$ of class C''' on which (2) and (8) hold. That S, in a Cartesian parametrization, say $z = z(x, y)$, is of class C'''' follows from the considerations in [4], § 14.

The variant of the theorem of Bonnet, proved in [3], shows that if ϕ is of class C', then there exists an $S: X = X(u, v)$ of class C'' on which (2) and (8) hold and which, by [4], § 14, is of class C''' in a suitable parametrization.

But if ϕ is only continuous, one cannot write down all of the equations in the linear total system, consisting of the derivation formulae of Gauss and Weingarten, for which the theorem of Bonnet and its variant supply a solution. In fact, these equations in the smooth cases of (2) and (8) are

$$(9)\quad X_{uu} = (X_u \cos \phi - X_v)\phi_u/\sin \phi, \qquad X_{vv} = (X_v \cos \phi - X_u)\phi_v/\sin \phi,$$

$$(10)\qquad X_{uv} = N \sin \phi$$

and, if $N = N(u, v)$ denotes the unit normal vector, $(X_u, X_v)/|(X_u, X_v)|$,

$$(11)\quad N_u = (X_u \cos \phi - X_v)/\sin \phi, \qquad N_v = (X_v \cos \phi - X_u)/\sin \phi,$$

while (9) involves partial derivatives of ϕ.

The existence assertion in (*) above will be proved by an approximation process. The uniqueness assertion will depend on the result of Part II, below.

3. Before beginning the proof, it can be remarked that the assertion (*) goes beyond the existence statements just mentioned; that is, there actually exist functions $\phi = \phi(u, v)$ satisfying (4), which are continuous but not of class C'. This can be proved by the existence proof of Picard for the hyperbolic differential equation (7). Let $\alpha = \alpha(u)$ and $\beta = \beta(v)$ be continuous functions for $|u| \leqq a$; $|v| \leqq b$, respectively, satisfying $\alpha(0) = \beta(0)$. Then there exists one and only one continuous function ϕ on the rectangle (6) satisfying (4) and $\phi(u, 0) = \alpha(u)$ and $\phi(0, v) = \beta(v)$. This follows by considering the successive approximations $\psi_0(u, v) = \alpha(u) + \beta(v) - \alpha(0)$ and

$$\psi_n(u, v) = \psi_0(u, v) + \int_0^u \int_0^v \sin \psi_{n-1}(u, v)\, du dv \ \text{ for } n = 1, 2, \cdots.$$

The standard calculation shows that $\phi = \lim \psi_n$, as $n \to \infty$, exists and is the desired function. Clearly, $\phi(u, v)$ is of class C' if and only if the given functions $\alpha(u)$, $\beta(v)$ are.

This proof also implies that if a ϕ is given, then there exists a sequence

of functions $\phi^1(u,v), \phi^2(u,v), \cdots$ on (6), which are smooth, satisfy (4) (that is, (7)) and which, as $n \to \infty$, tend uniformly to $\phi(u,v)$. In fact, it is sufficient to choose sequences $\alpha^n(u)$, $\beta^n(v)$ of polynomials, which satisfy $\alpha^n(0) = \beta^n(0)$ and which approximate uniformly the respective functions $\alpha(u) = \phi(u,0)$, $\beta(v) = \phi(0,v)$, and to let $\phi^n(u,v)$ be the unique solution of (7) belonging to the initial conditions $\phi^n(u,0) = \alpha^n(u)$, $\phi^n(0,v) = \beta^n(v)$.

4. In order to prove the existence statement in (*), let $\phi = \phi(u,v)$ be a given continuous function on (16) satisfying (4) and $0 < \phi < \pi$. Let $\phi^1, \phi^2, \cdots$ be a sequence of approximating smooth functions, described above. By virtue of uniform convergence, $0 < \phi^n(u,v) < \pi$ holds for all (u,v) on (6) and for all sufficiently large n. By discarding a finite number of the $\phi^1, \phi^2, \cdots$ (and renumbering the sequence), it can be supposed that this inequality holds for all n.

Let (2_n), (7_n), (8_n), $\cdots$ denote the relations (2), (7), (8), $\cdots$, respectively, in which ϕ is replaced by ϕ^n. By the theorem of Bonnet, there exists a unique smooth surface $S^n: X = X^n(u,v)$ for which (2_n), (8_n) hold and for which

$$(12_n) \qquad \begin{gathered} X^n(0,0) = 0, \qquad X_u{}^n(0,0) = (1,0,0), \\ X_v{}^n(0,0) = (\cos\phi^n(0,0), \sin\phi^n(0,0), 0). \end{gathered}$$

Note that $X^n(u,v)$ exists on the entire rectangle (6), since the total system involved, (9_n)-(11_n), is linear.

Let it be granted for the moment that *the sequences* $X_u{}^1, X_u{}^2, \cdots$ and $X_v{}^1, X_v{}^2, \cdots$ *are equicontinuous on* (6). These sequences are obviously bounded since (2_n) implies $|X_u{}^n| \equiv |X_v{}^n| \equiv 1$. Hence, by the theorem of Arzelà, it is possible to select a subsequence of the surfaces $X^1, X^2, \cdots$, which will again be denoted by $X^1, X^2, \cdots$, such that $X^1, X^2, \cdots$ and its sequence of first order partial derivatives converge uniformly on (6), as $n \to \infty$. Let $S: X = X(u,v)$ denote the limit surface.

Clearly, on the surface S, the squared element of arc-length is given by (2). Furthermore, if $N = N(u,v)$ is the unit normal vector, then N is of class C'. In fact, (11) holds as a consequence of (11_n) and the selection process above. In the same way, it is seen that $X_{uv} = X_{vu}$ exists, is continuous and satisfies (10).

The initial conditions (12) imply that the Jacobian of (1) does not vanish at $(u,v) = (0,0)$, so that S has a representation of the form $z = z(x,y)$, where $z = z(x,y)$ is of class C' in a vicinity of $(x,y) = (0,0)$.

But since $N = \pm(z_x, z_y, -1)/(1 + z_x^2 + z_y^2)^{\frac{1}{2}}$ is of class C' as a function of (u, v), and hence as a function of (x, y), it follows that $z = z(x, y)$ is of class C''. The relations (4) and (10) and the theorem of Gauss-Bonnet, as applied in [4], § 15, show that the curvature of S is identically -1. Finally, a point P of S and an asymptotic direction at P determine a unique asymptotic arc on S; [2], § 1. Hence, the asymptotic arcs on S^n, $u =$ const. and $v =$ Const., tend to those of S. Consequently, the parameter lines $u =$ const. and $v =$ Const. on S are the asymptotic arcs.

5. Thus, in order to complete the existence statement contained in (*), it remains to verify the hypothesis made in the last section, that the sequences $X_u^1, X_u^2, \cdots$ and $X_v^1, X_v^2, \cdots$ are equicontinuous, by virtue of the fact that $\phi^1, \phi^2, \cdots$ is. To this end, the differences $| X_u^n(u + h, v) - X_u^n(u, v) |$, $| X_u^n(u, v + h) - X_u^n(u, v) |$ and those belonging to X_v^n will be estimated in terms of those belonging to ϕ^n (and in terms of the lower bound for $\sin \phi^n$). Since n is fixed, the notation will be simplified by omitting n.

It will be sufficient to consider the differences belonging to X_u, as those belonging to X_v can be treated similarly. Let m satisfy

$$\sin \phi(u, v) \geqq m > 0 \text{ for all } (u, v) \text{ on } R. \tag{13}$$

If (u, v), $(u, v + h)$ are two points of R, it follows from (10) that

$$| X_u(u, v + h) - X_u(u, v) | \leqq | h | \tag{14}$$

since $| N | = 1$.

If $(u + h, v)$, (u, v) are two points of R, let Δf denote the difference $f(u + h, v) - f(u, v)$, where f is any (scalar or vector) function on R. The difference ΔX_u will be appraised by a method similar to that used in [4], § 13.

Let C be a (fixed) number to be specified below and let $\epsilon > 0$ be any number satisfying

$$2C\epsilon < m^3/4, \tag{15}$$

where m satisfies (13). Corresponding to ϵ, there exists a number $\delta = \delta_\epsilon > 0$ with the property that

$$| \Delta\phi | < \epsilon \text{ if } | h | < \delta \tag{16}$$

for all (u, v), $(u + h, v)$ in R. For convenience later, it can be supposed that δ is chosen so small that

$$2\delta < \epsilon m. \tag{17}$$

It will be shown that

$$(18) \qquad |\Delta X_u| < 8C\epsilon/m^2 \text{ if } |h| < \delta$$

for all (u, v), $(u+h, v)$ on R. If (18) is verified, then it follows from the definition (16), (17) of δ, that the existence statement in (*) is proved.

Let

$$(19) \qquad A_1 = X_u(u,v) + \tfrac{1}{2}\Delta X_u, \qquad A_2 = X_v(u,v), \qquad A_3 = N(u,v).$$

If these three vectors are linearly independent, that is, if

$$(20) \qquad \det(A_1, A_2, A_3) \neq 0,$$

then there exist three unique numbers $\alpha_i = \alpha_i(u, v, h)$, where $i = 1, 2, 3$, such that

$$(21) \qquad \Delta X_u = \alpha_1 A_1 + \alpha_2 A_2 + \alpha_3 A_3.$$

If (21) is multiplied scalarly by A_1, A_2, A_3, respectively, there results a system of linear equations for α_1, α_2, α_3. It will be seen that these equations are

$$(22_1) \quad 0 = \alpha_1(A_1 \cdot A_1) + \alpha_2(A_2 \cdot A_1) + \alpha_3(A_3 \cdot A_1),$$

$$(22_2) \quad \Delta \cos\phi - X_u(u+h, v) \cdot \Delta X_v = \alpha_1(A_1 \cdot A_2) + \alpha_2(A_2 \cdot A_2) + \alpha_3(A_3 \cdot A_2),$$

$$(22_3) \quad -X_u(u+h, v) \cdot \Delta N = \alpha_1(A_1 \cdot A_3) + \alpha_2(A_2 \cdot A_3) + \alpha_3(A_3 \cdot A_3).$$

The right-hand sides of these equations are obvious. The left-hand sides can be verified as follows: the left-hand side of (22_1) is $\Delta X_u \cdot A_1 = \Delta(|X_u|^2) = \Delta(1) = 0$; the left-hand side of (22_2) is

$$\begin{aligned} \Delta X_u \cdot A_2 &= \Delta(X_u \cdot A_2) - X_u(u+h, v) \cdot \Delta A_2 \\ &= \Delta \cos\phi - X_u(u+h, v) \cdot \Delta X_v; \end{aligned}$$

finally, the left hand side of (22_3) is

$$\begin{aligned} \Delta X_u \cdot A_3 &= \Delta(X_u \cdot A_3) - X_u(u+h, v) \cdot \Delta A_3 \\ &= \Delta(0) - X_u(u+h, v) \cdot \Delta N. \end{aligned}$$

It is clear from (10), (11) and (13) that

$$(23) \qquad |\Delta X_v| \leqq |h| \text{ and } |\Delta N| \leqq 2|h|/m.$$

Hence, the terms on the left-hand sides of (22_1)-(22_3) are majorized by $|\Delta\phi| + 2|h|/m$. Since the elements of the matrix $(A_i \cdot A_j)$ are majorized by $|A_i||A_j| < 2 \cdot 2 = 4$, it follows that the two-rowed minors are majorized by 32. Hence, if (20) holds,

$$|\alpha_i| \leqq 96(|\Delta\phi| + 2|h|/m)\det^{-2}(A_1, A_2, A_3),$$

since $\det^2(A_1, A_2, A_3) = \det(A_i \cdot A_j)$. Hence, (18) implies

$$(24) \qquad |\Delta X_u| \det{}^2(A_1, A_2, A_3) \leqq C(|\Delta\phi| + 2|h|/m),$$

if C is sufficiently large, say $C = 576$. Hence, by (16) and (17),

$$(25) \qquad |\Delta X_u| \det{}^2(A_1, A_2, A_3) \leqq C(\epsilon + 2\delta/m) \leqq 2C\epsilon$$

if $|h| < \delta$.

It will now be shown that

$$(26) \qquad |\Delta X_u| < m \text{ if } |h| < \delta$$

for all (u, v), $(u + h, v)$ in R. By the definition (19) of A_1, A_2, it is seen that the vector product (A_1, A_2) is $N \sin\phi + \frac{1}{2}(\Delta X_u, X_v)$; so that $\det(A_1, A_2, A_3)$ is $\sin\phi + \frac{1}{2}\det(\Delta X_u, X_v, N)$. Consequently,

$$(27) \qquad |\det(A_1, A_2, A_3)| \geqq m - \tfrac{1}{2}|\Delta X_u|.$$

Suppose, if possible, that the first inequality in (26) is violated for a pair of points $(u + h, v)$, (u, v) of R, where $|h| < \delta$. Clearly, it can then be supposed that the first inequality sign in (26) can be replaced by equality if $(u + h, v)$, (u, v) are suitable chosen. By (27), $|\det(A_1, A_2, A_3)| \geqq \frac{1}{2}m > 0$; and so, (25) implies $|\Delta X_u| \leqq (2C\epsilon)/(\frac{1}{2}m)^2$. By (15), this means $|\Delta X_u| < m$, which contradicts the assumption $|\Delta X_u| = m$. Consequently, (26) holds.

Hence, (25) and (27) imply (18) for all $(u + h, v)$, (u, v) in R. This completes the proof of the existence statement in (*).

6. There remains to prove the uniqueness statement contained in the assertion (*). Let $S_1 : z = z_1(x_1, y_1)$ and $S_2 : z = z_2(x_2, y_2)$ be two surfaces satisfying the statement of the first paragraph of § 1, where the function $\phi = \phi(u, v)$ for both surfaces is the given ϕ. It will be verified that S_1 and S_2 are identical, up to a Euclidean movement.

In the (x_1, y_1)-parametrization of S_1, the squared element of arc-length on S_1 is

$$(28) \qquad ds^2 = (1 + p^2)dx_1^2 + 2pq\,dx_1dy_1 + (1 + q^2)dy_1^2,$$

where $p = \partial z_1/\partial x_1$ and $q = \partial z_1/\partial y_1$. The surfaces S_1, S_2 have parametrizations $X = X_1(u, v)$, $X = X_2(u, v)$ of class C', in which (2) holds with the same given ϕ. Let $x = x_2(x_1, y_1)$, $y = y_2(x_1, y_1)$ be the result of the transformations $(x_2, y_2) \to (u, v) \to (x_1, y_1)$. Thus S_2 has a parametrization of the form

$$(29) \qquad S_2 : x = x_2(x_1, y_1),\ y = y_2(x_1, y_1),\ z = z_2(x_2(x_1, y_1), y_2(x_1, y_1)),$$

which is of class C'. In the (x_1, y_1)-parametrization, it follows from the standard transformation rule for ds^2, that (28) holds on S_1. Since the coefficients in (28) are of class C_1, it follows from (**) in Part II below that the transformation $(x_1, y_1) \rightarrow (x_2, y_2)$ is of class C''; and so the (x_1, y_1)-parametrization of both S_1 and S_2 are of class C''.

It can easily be verified that, if the second fundamental form in the (u, v)-parameters are calculated formally by the standard transformation rule (or equivalently by, $h_{11} = -X_u \cdot N_u$, $h_{12} = h_{21} = -X_u \cdot N_v$, $h_{22} = -X_v \cdot N_v$) for both S_1 and S_2, then $h_{ik}du^i du^k = \pm 2 \sin \phi \, dudv$; and it can, therefore, be supposed that (8) holds. Hence, in terms of the (x_1, y_1)-parameters, S_1 and S_2 have the same second fundamental form.

Thus, the surfaces $S_1 : z = z_1(x_1, y_1)$ and (29) are of class C'' and have the same first and second fundamental forms. It follows from the uniqueness theorem in [3], § 2 that S_1 and S_2 are identical, up to a Euclidean movement. This completes the proof of the assertion (*).

Part II.

7. In this part, the following theorem will be proved:

(**) *Let* $(g_{ik}) = (g_{ik}(u^1, u^2))$, *where* $i, k = 1, 2$, *be a positive-definite symmetric matrix of class* C' *in a vicinity of* $(u^1, u^2) = (0, 0)$. *Let*

$$u^i = u^i(U^1, U^2), \text{ where } i = 1, 2, \qquad (u^i(0,0) = 0), \tag{30}$$

be a transformation of class C' *in a vicinity of* $(U^1, U^2) = (0, 0)$ *with a non-vanishing Jacobian carrying*

$$ds^2 = g_{ik}du^i du^k \tag{31}$$

into

$$ds^2 = G_{ik}dU^i dU^k. \tag{32}$$

If (30) *has the property that the symmetric matrix* $(G_{ik}) = (G_{ik}(U^1, U^2))$ *is of class* C', *then the transformation* (30) *is of class* C''.

This theorem is a generalization of the lemma of [4], § 12. In addition to the application of (**) in § 6, above, several other consequences of (**) can be mentioned:

(i) From Wintner's discussion [9] of the notion of isometry in differential geometry, the usefulness of (**) is at once apparent. In fact, (**) and its analogues (cf. the remark at the end of this section) show that if two surfaces of class C^n, where $n \geqq 1$, are isometric (by virtue of a trans-

formation of class C^1), then they are isometric by virtue of a transformation of class C^n. This answers the question raised by Wintner [9], end of § 4, concerning the interpretation of certain standard theorems in the theory of surfaces.

(ii) If the coefficients g_{ik} of (31) are of class C' and (31) possesses a curvature $K = K(u, v)$ in the sense of Weyl [7] (cf., [6]) and K is of class C', then it follows from [1], § 6 and from (**) that it is possible to introduce local geodesic parallel coordinates (U^1, U^2) and that the transformation (30) is of class C''. (For contrast, it can be mentioned that such a transformation (30) need not exist if a bounded curvature K does not exist (cf. [1], § 2); the transformation (30) exists and is of class C' when a bounded curvature K does exist (cf. the proofs of Theorems 1 and 2 in [1] and (III) in [5]), but need not be of class C'' if K is not of class C' (cf. [2]).

(iii) If the surface S has a parametrization of class C'' in terms of some parameters, say (u^1, u^2), and if (30) is a transformation of class C' such that the first fundamental form $G_{ik}dU^idU^k$ and formal (cf. § 6 above) second fundamental form $H_{ik}dU^idU^k$ are of class C', then, by (**) and the considerations of [4], § 14, the surface S is of class C''' in a suitable parametrization, say $z = z(x, y)$.

To illustrate the principle (iii), let S possess a negative curvature of class C', then it is possible to introduce (locally) the asymptotic lines as coordinate curves $U^1 = \text{const.}$, $U^2 = \text{Const.}$, and the transformation (30) is of class C', [4], § 6. If the coefficients in the resulting squared element of arc-length (32) are of class C', then the above principle is applicable, since $H_{11} = H_{22} = 0$ and $H_{12}{}^2 = -\det(G_{ik})/K > 0$ are of class C'. Thus, the asymptotic line parametrization of a surface (of negative curvature K of class C') cannot be of class C'' unless the surface has some parametrization of class C'''. The case $K \equiv -1$ of this illustration is proved in [4] by using the lemma, [4], § 12, mentioned above.

As another illustration of (iii), let S possess distinct principle curvatures of class C', then it is possible to introduce (locally) the lines of curvature as coordinate curves $U^1 = \text{const.}$, $U^2 = \text{Const.}$, and the transformation (30) is of class C', [4], § 17. If the coefficients in the resulting squared element of arc-length (32) are of class C', then the above principle is applicable, since $H_{12} = 0$ and H_{11}/G_{11}, H_{22}/G_{22}, being the respective principal curvatures, are of class C'. Thus, the line of curvatures parametrization of a surface (having distinct principal curvatures of class C') cannot be of class C'' unless the surface has some parametrization of class C'''.

Whether or not (**) is true if the symmetric matrix (g_{ik}) is either n by n and non-singular (instead of being 2 by 2 and positive-definite) will remain undecided. (The proof of (**) and this generalization would follow, for instance, if it could be shown that the transformation (30) can be approximated by smooth transformations in such a way that the Christoffel symbols belonging to the forms corresponding to (32) tend uniformly to those belonging to (32)). An indication of the truth in the n by n, positive-definite case is given by the results in [8].

The theorem (**) is obviously false if (g_{ik}) is singular. For consider the cases $u = U$, $v = v(U, V)$ and $ds^2 = du^2 + 0 \cdot dudv + 0 \cdot dv^2$ of (30) and (31), respectively. The corresponding form (32) is $ds^2 = dU^2 + 0 \cdot dUdV + 0 \cdot dV^2$, while $v(U, V)$ need not be of class C''.

The replacement of C', C'' by C^m, C^{m+1}, respectively, where $m > 1$, in (**) leads to a correct but easily proved theorem, in view of the transformation rule for Christoffel symbols. The difficulty in the proof of (**) is to establish the validity of this rule when $m = 1$.

8. *Proof of* (**). Since the coefficient functions of both (31) and (32) are of class C', the Christoffel symbols of the second kind, $\gamma_{jk}{}^i = \gamma_{jk}{}^i(u^1, u^2)$ and $\Gamma_{jk}{}^i = \Gamma_{jk}{}^i(U^1, U^2)$, belonging to (31) and (32), respectively, exist and are continuous. The two systems of differential equations

$$(33) \qquad u^{i\prime\prime} + \gamma_{jk}{}^i u^{j\prime} u^{k\prime} = 0, \text{ where } i = 1, 2, \qquad (' = d/ds),$$

and

$$(34) \qquad U^{i\prime\prime} + \Gamma_{jk}{}^i U^{j\prime} U^{k\prime} = 0, \text{ where } i = 1, 2, \qquad (' = d/ds),$$

define the geodesics belonging to (31) and (32), respectively. Formally, it cannot, however, be verified that the solutions of (34) correspond, by virtue of (30), to solutions of (33), and conversely. (Of course, such a verification could be made by considering the geodesics as extremals (but not "minimizing" curves, cf. (II) in [5]) of a calculus of variations problem.)

If a point (u^1, u^2) and/or (U^1, U^2) is sufficiently near $(0, 0)$, then there exist arcs minimizing the distance, in the metric (31) and/or (32), between $(0, 0)$ and that point (Hilbert). All such arcs are solutions of (33) and/or (34); cf. the proof of (I) in [5]. If $u^i = u^i(s)$ and $U^i = U^i(s)$ is such a geodesic, where the initial conditions are chosen so that (30) gives

$$(35) \qquad u^{i\prime} = u_j{}^i U^{j\prime}, \text{ where } u_j{}^i = \partial u^i / \partial U^j,$$

then both $u^i(s)$ and $U^i(s)$ are of class C''. (Incidentally, not all geodesics minimize the distance between sufficiently close points on them; cf. (II) in [5]).

Corresponding to a geodesic $u^i = u^i(s)$, $U^i = U^i(s)$, related by (30) and (35), consider the linear differential equations

$$(36) \qquad w^{i\prime} + \gamma_{jk}{}^i w^j u^{k\prime} = 0, \text{ where } i = 1, 2$$

and

$$(37) \qquad W^{i\prime} + \Gamma_{jk}{}^i W^j U^{k\prime} = 0, \text{ where } i = 1, 2,$$

with $\gamma_{jk}{}^i = \gamma_{jk}{}^i(u^1(s), u^2(s))$ and $\Gamma_{jk}{}^i = \Gamma_{jk}{}^i(U^1(s), U^2(s))$ in (36) and (37), respectively. A solution $W^i = W^i(s)$ of (37) is a field of parallel vectors along the geodesic $U^i = U^i(s)$. Hence, the scalar products $G_{ik} W^i W^k$, $G_{ik} W^i U^{k\prime}$, which are first integrals of (37), are independent of s. But scalar products are invariant under transformations of class C'; and so

$$(38) \qquad w^i = u_j{}^i W^j, \text{ where } i = 1, 2,$$

with $u_j{}^i = u_j{}^i(U^1(s), U^2(s))$, is a solution of (36) and, in particular, is of class C'.

It is clear from (34) and the fact that $G_{ik} U^{i\prime} U^{k\prime} = 1$ along solutions of (34), while (32) is positive-definite, that $| U^{i\prime\prime} | \leqq \text{Const.}$, whenever $(U^1(s), U^2(s))$ is sufficiently near $(0, 0)$, where Const. is independent of the geodesic $U^i = U^i(s)$. This implies that if J_Δ is a geodesic arc $U^i = U^i(s)$ $= U^i(s, \Delta U^1)$ joining $(0, 0)$ and $(\Delta U^1, 0)$, where $| \Delta U^1 | \neq 0$ is sufficiently small, then $(U^{1\prime}(s), U^{2\prime}(s))$ tends, uniformly on J_Δ, to $(G_{11}^{-\frac{1}{2}}(0, 0), 0)$, as $\Delta U^1 \to 0$. Similarly, if $W^i = W^i(s) = W^i(s, \Delta U^1)$ is a field of parallel vectors on J_Δ determined by initial conditions $(W_0{}^1, W_0{}^2)$ at $(0, 0)$, which are independent of ΔU^1, then $(W^1(s), W^2(s))$ tends, uniformly on J_Δ, to $(W_0{}^1, W_0{}^2)$, as $\Delta U \to 0$. Finally, if Δs is the length of J_Δ, then $\Delta s / \Delta U^1$ $\to G_{11}^{\frac{1}{2}}(0, 0)$, as $\Delta U^1 \to 0$.

Let the geodesic arc J_Δ be chosen so that $u^i = u^i(s)$, determined from (30), is a solution of (33). Let f_0, where f is $w^i, u_j{}^i, \cdots$, denote the value of f at $(0, 0)$ and let Δf denote the difference between the values of f at the points $(\Delta U^1, 0)$ and $(0, 0)$. Then, by (38),

$$(39) \qquad \Delta w^i = (\Delta u_j{}^i) W_0{}^j + u_j{}^i \Delta W^j.$$

By (37) and the mean-value theorem of differential calculus,

$$\Delta W^j = -(\Gamma_{nm}{}^j W^n U^{m\prime}) \Delta s,$$

where the coefficient of Δs is evaluated at some intermediary point on J_Δ (depending on j and ΔU^1). By the remarks of the last paragraph, $\Delta W^j / \Delta U^1$ tends to $-\Gamma_{n10}{}^j W_0{}^n$, as $\Delta U^1 \to 0$. Similarly, $\Delta w^i / \Delta U^1$ tends, as $\Delta U^1 \to 0$, to $-\gamma_{km0}{}^i w_0{}^k u_0{}^{m\prime} G_{110}{}^{-\frac{1}{2}}$. By (35) and (38), the latter limit is $-\gamma_{km0}{}^i u_{n0}{}^k u_{10}{}^m W_0{}^n$.

Hence (39) implies that $(\Delta u_j{}^i/\Delta U^1)W_0{}^j$ tend to a limit, as $\Delta U^1 \to 0$. Since $W_0{}^j$ is arbitrary, it follows that, as $\Delta U^1 \to 0$, the quotient

$$\Delta u_j{}^i/\Delta U^1 = \{u_j{}^i(\Delta U^1, 0) - u_j{}^i(0,0)\}/\Delta U^1$$

tends to the limit $\Gamma_{j1}{}^n u_n{}^i - \gamma_{km}{}^i u_j{}^k u_1{}^m$, evaluated at $(0,0)$. Since ΔU^1 can be replaced by ΔU^2 in this argument, and $(0,0)$ by any point (U^1, U^2) sufficiently near $(0,0)$, it follows that the second order partial derivatives of the functions in (30) exist and can be calculated from the standard transformation rule for Christoffel symbols of the second kind. Consequently, (**) is proved.

PARIS, FRANCE.

REFERENCES.

[1] P. Hartman, "On the local uniqueness of geodesics," *American Journal of Mathematics*, vol. 72 (1950), pp. 723-730.

[2] ———, "On geodesic coordinates," *ibid.*, vol. 73 (1951), pp, 949-954.

[3] ——— and A. Wintner, "On the fundamental equations of differential geometry," *ibid.*, vol. 72 (1950), pp. 757-774.

[4] ——— and A. Wintner, "On the asymptotic curves of a surface," *ibid.*, vol. 73 (1951), pp. 149-172.

[5] ——— and A. Wintner, "On the problems of geodesics in the small," *ibid.*, vol. 73 (1951), pp. 132-148.

[6] E. R. van Kampen, "The theorems of Gauss-Bonnet and Stokes," *ibid.*, vol. 60 (1938), pp. 129-138.

[7] H. Weyl, "Ueber die Bestimmung einer geschlossenen konvexen Fläche durch ihr Linienelement," *Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich*, vol. 61 (1915), pp. 40-72.

[8] A. Wintner, "On Riemann metrics of constant curvature," *American Journal of Mathematics*, vol. 73 (1951), pp. 569-575.

[9] ———, "On isometric surfaces," *ibid.*, vol. 74 (1952), pp. 198-214.

ON THE KINEMATIC FORMULA IN THE EUCLIDEAN SPACE OF N DIMENSIONS.*

By SHIING-SHEN CHERN.

Introduction. The idea of considering the kinematic density in problems of geometrical probability was originated by Poincaré. It was further exploited by L. A. Santalò and W. Blaschke in their work on integral geometry [1], culminating in the following theorem:

Let Σ_0, Σ_1 be two closed surfaces in space, which are twice differentiable, and let D_0, D_1 be the domains bounded by them. Let V_i, $\chi_i = K_i/4\pi$ be the volume and Euler characteristic of D_i and let A_i, M_i be the area and the integral of mean curvature of Σ_i, $i = 0, 1$. Suppose Σ_0 fixed and Σ_1 moving. Then the integral of $K(D_0 \cdot D_1) = 4\pi\chi(D_0 \cdot D_1)$ over the kinematic density of Σ_1 is given by the formula

$$\int K(D_0 \cdot D_1)\dot{\Sigma}_1 = 8\pi^2(V_0K_1 + A_0M_1 + M_0A_1 + K_0V_1). \tag{1}$$

This formula includes most formulas in Euclidean integral geometry as special or limiting cases. The purpose of this paper is to apply E. Cartan's method of moving frames and to derive the generalization of this formula in an Euclidean space of n dimensions. By doing this, we hope that some insight can be gained on integral geometry in a general homogeneous space. Moreover, one of the ideas introduced, the consideration of measures in spaces which are now called fiber bundles, will most likely find further applications. The main procedures of our proof have been given in a previous note [2].

We consider a compact orientable hypersurface Σ, twice differentiably imbedded in an Euclidean space E of n $(\geqq 2)$ dimensions. At a point P of Σ there are $n-1$ principal curvatures κ_α, $\alpha = 1, \cdots, n-1$, whose i-th elementary symmetric function we shall denote by S_i, $i = 0, \cdots, n-1$, where $S_0 = 1$ by definition. Let dA be the element of area of Σ, and let

$$M_i = \int_\Sigma S_i dA \Big/ \binom{n-1}{i}, \qquad i = 0, 1, \cdots, n-1. \tag{2}$$

These M_i are integro-differential invariants of Σ. In particular, M_0 is the

* Received May 10, 1951.

area and M_{n-1} is a numerical multiple of the degree of mapping of Σ into the unit hypersphere defined by the field of normals.

Take now two such hypersurfaces Σ_0, Σ_1, whose invariants we distinguish by superscripts. The volume of the domain D_i bounded by Σ_i we denote by V_i, $i = 0, 1$. Let Σ_0 be fixed and Σ_1 be moving, and let $\dot{\Sigma}_1$ be the kinematic density of Σ_1. We suppose our hypersurfaces to be such that for all positions of Σ_1 the intersection $D_0 \cdot D_1$ has a finite number of components. Then the Euler-Poincaré characteristic $\chi(D_0 \cdot D_1)$ is well defined. If I_{n-1} denotes the area of the unit hypersphere in E and if

$$J_n = I_1 I_2 \cdots I_{n-1}, \tag{3}$$

the *kinematic formula* in E is

$$\int K(D_0 \cdot D_1)\dot{\Sigma}_1 = J_n\{M_{n-1}^{(0)}V_1 + M_{n-1}^{(1)}V_0 + \frac{1}{n}\sum_{k=0}^{n-2}\binom{n}{k+1} M_k^{(0)}M_{n-2-k}^{(1)}\}, \tag{4}$$

where

$$K(D_0 \cdot D_1) = I_{n-1}\chi(D_0 \cdot D_1). \tag{5}$$

For $n = 3$ this reduces to the formula (1). The formula for $n = 4$ is

$$\int K(D_0 \cdot D_1)\dot{\Sigma}_1 = 16\pi^4(M_3^{(0)}V_1 + M_3^{(1)}V_0 + M_0^{(0)}M_2^{(1)} + M_0^{(1)}M_2^{(0)} + \tfrac{3}{2}M_1^{(0)}M_1^{(1)}). \tag{6}$$

1. Measures in spaces associated with a Riemann manifold. We shall first review a few notions in Riemannian geometry, in a form which will be useful for our later purpose.

Let M be an orientable Riemann manifold of class $\geqq 3$ and dimension n. Associated with M are the spaces B_h $(h = 1, \cdots, n)$ formed by the elements $Pe_1 \cdots e_h$, each of which consists of a point P of M and an ordered set of h mutually perpendicular tangent unit vectors $e_1, \cdots, e_h$ at P. When $h = n$, such an element will be called a frame. In the current terminology B_n is a principal fiber bundle over M with the rotation group as structural group and B_h are the associated bundles [3]. We shall introduce a measure in B_h. Since B_h is clearly an orientable differentiable manifold, this can be done by defining an exterior differential form of degree $\frac{1}{2}(h+1)(2n-h)$ ($=$ dim of B_h).

There is a natural mapping $\psi_h : B_n \to B_h$ defined by taking as the image of $Pe_1 \cdots e_n$ the element $Pe_1 \cdots e_h$. It induces a dual homomorphism of the differential forms of B_h into those of B_n. This process has in a sense a converse. In fact, let

$$e_r = \sum_s u_{rs}e^*_s, \qquad h+1 \leqq r, s \leqq n \tag{7}$$

be a rotation of the last $n-h$ vectors. A differential form of B_n which is invariant under the action of (7) can be regarded as a form of B_n.

The well-known parallelism of Levi-Civita can be interpreted as defining a set of $n(n+1)/2$ linearly independent Pfaffian forms in B_n, which we shall denote by ω_i, $\omega_{ij}(=-\omega_{ji})$, $1 \leqq i, j \leqq n$. To give it a brief description [4] we start from the following useful lemma on exterior forms: *Let ω_i be linearly independent Pfaffian forms, and let $\pi_{ij}=-\pi_{ji}$ be Pfaffian forms such that*[1]

$$\sum_j \omega_j \wedge \pi_{ji} = 0. \tag{8}$$

Then $\pi_{ij}=0$. In fact, it follows from (8) that

$$\pi_{ji} = \sum_{k=1}^{n} a_{jik}\omega_k.$$

Then a_{jik} is skew-symmetric in its first two indices, because the π_{ji} are, and is symmetric in its last two indices, on account of (8). Therefore $a_{jik}=0$ or $\pi_{ji}=0$.

For geometric reasons we denote by dP the identity mapping in the tangent space at P, which maps every tangent vector into itself. Then dP can be written in the form

$$dP = \sum_i \omega_i \otimes e_i, \tag{9}$$

where the multiplication is tensor product, and the ω_i are Pfaffian forms in B_n and are linearly independent. The fundamental theorem on local Riemannian geometry asserts that there exists a uniquely determined set of Pfaffian forms ω_i, ω_{ij} in B_n, linearly independent, which satisfy (9) and

$$d\omega_i = \sum_j \omega_j \wedge \omega_{ji}. \tag{10}$$

In fact, the uniqueness follows from the above lemma.

For our purpose we shall study the effect of the rotation (7) on these forms. Denote the new forms by the same symbols with asterisks. Clearly we have

$$\omega_\alpha^* = \omega_\alpha, \qquad \omega_r^* = \sum_s u_{sr}\omega_s, \qquad 1 \leqq \alpha \leqq h,\ h+1 \leqq r, s \leqq n. \tag{11}$$

Taking the exterior derivatives of both sides of these equations and making use of (10), we get

---

[1] We shall, following Bourbaki, use wedge product to denote exterior multiplication. It will sometimes be dropped, when the meaning is clear. Parentheses will be used to denote ordinary products of differential forms.

$$\begin{aligned}(12)\qquad &\sum_\beta \omega_\beta \wedge (\omega_{\beta\alpha}^* - \omega_{\beta\alpha}) + \sum_r \omega_r^* \wedge (\omega_{r\alpha}^* - \sum_s u_{sr}\omega_{s\alpha}) = 0,\\ &\sum_\alpha \omega_\alpha \wedge (\omega_{\alpha r}^* - \sum_s u_{sr}\omega_{\alpha s}) + \sum_s \omega_s^* \wedge \phi_{sr}^* = 0,\end{aligned}$$

where ϕ_{sr}^* are Pfaffian forms skew-symmetric in the indices s, r. The system of equations (12) is of the same form as (8), and the above lemma is then applicable. It follows that

$$(13)\qquad \omega_{\beta\alpha}^* = \omega_{\beta\alpha}, \qquad \omega_{\alpha r}^* = \sum_s u_{sr}\omega_{\alpha s}.$$

If we put

$$(14)\qquad \Omega_\alpha = \prod_r \omega_{\alpha r},$$

we see from (13) that Ω_α is invariant under the action of (7). The same is therefore true of the form

$$(15)\qquad L_{n,h} = \prod_\alpha \Omega_\alpha \prod_{\alpha<\beta} \omega_{\alpha\beta} \prod_i \omega_i.$$

This form is clearly not identically zero, and we define it to be the density in B_h. It gives rise to a measure in B_h.

2. Differential geometry of a submanifold in Euclidean space. As a further preparation we need some notions on the geometry of a hypersurface in Euclidean space. As no additional complication is involved, we develop them for a submanifold V of p dimensions, which is twice differentiably imbedded in E. We agree in this section on the following ranges of indices:

$$(16)\qquad 1 \leqq \alpha, \beta, \gamma \leqq p, \qquad p+1 \leqq r, s, t \leqq n, \qquad 1 \leqq i, j, k \leqq n.$$

Since E is a Riemann manifold, the discussions of the last section are valid. In this case B_n is naturally homeomorphic to the group of proper motions in E. To study V we consider the submanifold of B_n characterized by the conditions that $P \varepsilon V$ and that the e_α are tangent vectors to V at P. If we denote by the same notation the forms on this submanifold induced by the identity mapping, we have

$$(17)\qquad \omega_r = 0.$$

From (10) it follows that

$$d\omega_r = \sum_\alpha \omega_\alpha \wedge \omega_{\alpha r} = 0.$$

Since the ω_α are linearly independent, we have

$$(18)\qquad \omega_{r\alpha} = \sum_\beta A_{r\alpha\beta}\omega_\beta$$

where the $A_{r\alpha\beta}$ are symmetric in α, β:

(19) $$A_{r\alpha\beta} = A_{r\beta\alpha}.$$

From these Pfaffian forms it is possible to construct some significant "ordinary" quadratic differential forms. The first is a set

(20) $$\Phi_r = \sum_{\alpha}(\omega_{r\alpha}\omega_\alpha) - \sum_{\alpha,\beta} A_{r\alpha\beta}(\omega_\alpha\omega_\beta),$$

which generalizes the second fundamental form in ordinary surface theory. The second is

(21) $$\Psi = \sum_{r,\alpha}(\omega_{r\alpha})^2 - \sum_{r,\alpha,\beta,\gamma} A_{r\alpha\beta}A_{r\alpha\gamma}(\omega_\beta\omega_\gamma),$$

generalizing the third fundamental form. The latter seems to deserve some attention. However, so far as the writer is aware, it has not been considered in the literature.

For a hypersurface we have $p = n - 1$, and we shall write Φ, $A_{\alpha\beta}$ for Φ_n, $A_{n\alpha\beta}$ respectively. The $n-1$ roots of the characteristic equation

(22) $$| A_{\alpha\beta} - \kappa\delta_{\alpha\beta} | = 0$$

are called the principal curvatures.

In the case of the Euclidean space E we can also write ω_i, ω_{ij} as scalar products, thus:

(23) $$\omega_i = dP \cdot e_i, \qquad \omega_{ij} = de_i \cdot e_j.$$

3. A formula on densities. The situation we are going to consider consists of two hypersurfaces Σ_0, Σ_1 in E, with Σ_0 fixed and Σ_1 moving, which intersect in a manifold V^{n-2} of dimension $n-2$, such that at a point of V^{n-2} the normals to Σ_0, Σ_1 never coincide. We denote by ϕ, $\phi \neq 0, \pi$, the angle between these normals and by $\dot{\Sigma}_1$ the kinematic density of Σ_1. An $(n-2)$-frame on V^{n-2} has a density on each of V^{n-2}, Σ_0, Σ_1, to be denoted by L_V, $\dot{L}_0$, L_1 respectively. Our formula to be proved can be written

(24) $$\dot{L}_V\dot{\Sigma}_1 = \sin^{n-1}\phi \dot{L}_0\dot{L}_1\dot{\phi}.$$

Throughout this section we shall agree on the following ranges of indices:

(25) $$1 \leqq i, j, k \leqq n, \qquad 1 \leqq \alpha, \beta \leqq n-2, \qquad 1 \leqq A, B \leqq n-1.$$

Let $O\mathfrak{a}_1 \cdots \mathfrak{a}_n$ be the fixed frame and $O'\mathfrak{a}'_1 \cdots \mathfrak{a}'_n$ the moving frame. For a given relative position between $O\mathfrak{a}_1 \cdots \mathfrak{a}_n$ and $O'\mathfrak{a}'_1 \cdots \mathfrak{a}'_n$ let $Pe_1 \cdots e_{n-2}$ be an $(n-2)$-frame on V^{n-2}. We complement this into a frame $Pe_1 \cdots e_n$ such that e_n is normal to Σ_0 and also into a frame $P'e'_1 \cdots e'_n$

such that e'_n is normal to Σ_1 at P, and $P' = P$, $e'_\alpha = e_\alpha$. Between e_{n-1}, e_n, e'_{n-1}, e'_n we have then the relations

$$e'_{n-1} = \cos\phi e_{n-1} + \sin\phi e_n, \qquad e'_n = -\sin\phi e_{n-1} + \cos\phi e_n. \tag{26}$$

From this we derive the following useful relation

$$de'_{n-1}\cdot e'_n = d\phi + de_{n-1}\cdot e_n. \tag{27}$$

Let us now express the relations between the frames so introduced by the equations

$$\begin{aligned} P &= O + \sum_i x_i a_i, & e_i &= \sum_k u_{ik} a_k, \\ P' &= O' + \sum_i x'_i a'_i, & e'_i &= \sum_k u'_{ik} a'_k. \end{aligned} \tag{28}$$

We shall denote the differentiation by d' when $O'a'_1\cdots a'_n$ is regarded as fixed. In other words, d' is differentiation relative to the moving frame. Then we have, from (28),

$$dO' = dP - d'P - \sum_i x'_i da'_i. \tag{29}$$

It follows that, on neglecting terms in da'_i,

$$\begin{aligned} \prod_\alpha(dP\cdot e_\alpha)\prod_i(dO'\cdot a'_i) &\equiv \prod_\alpha(dP\cdot e_\alpha)\prod_i(dP\cdot a'_i - d'P\cdot a'_i) \\ &\equiv \prod_\alpha(dP\cdot e_\alpha)\prod_i(dP\cdot e'_i - d'P\cdot e'_i) \\ &\equiv \pm\prod_\alpha(dP\cdot e_\alpha)(d'P\cdot e'_\alpha)(dP\cdot e'_{n-1} - d'P\cdot e'_{n-1})(dP\cdot e'_n - d'P\cdot e'_n) \\ &\equiv \pm\sin\phi\prod_A(dP\cdot e_A)(d'P\cdot e'_A). \end{aligned} \tag{30}$$

These are to be taken as congruences mod da'_i. In particular, the last step follows from the fact that e'_n is normal to Σ_1 at P and that the product of n factors involving dP is zero, because the locus of P is a hypersurface Σ_0.

In order to get a further reduction of the left-hand side of (24) we start from the formula

$$de'_i - d'e'_i = \sum_k u'_{ik} da'_k. \tag{31}$$

From the invariance of the kinematic density under a rotation it follows that

$$\prod_{i<j}(da'_i\cdot a'_j) = \prod_{i<j}((de'_i - d'e'_i)\cdot e'_j).$$

Then we have

$$
\begin{aligned}
\prod_{\alpha<\beta}(de_\alpha\cdot e_\beta)\prod_{i<j}(da'_i\cdot a'_j) &= \prod_{\alpha<\beta}(de_\alpha\cdot e_\beta)\prod_{i<j}(de'_i - d'e'_i)\cdot e'_j \\
&= \prod_{\alpha<\beta}(de_\alpha\cdot e_\beta)\prod_{\alpha<i}((de_\alpha - d'e_\alpha)\cdot e'_i)\,((de'_{n-1} - d'e'_{n-1})\cdot e'_n) \\
&= \prod_{\alpha<\beta}(de_\alpha\cdot e_\beta)\prod_{\alpha<\beta}((de_\alpha - d'e_\alpha)\cdot e'_\beta) \\
&\qquad \wedge\prod_\alpha\{(de_\alpha\cdot e'_{n-1} - d'e_\alpha\cdot e'_{n-1})(de_\alpha\cdot e'_n - d'e_\alpha\cdot e'_n)\} \\
&\qquad \wedge(de'_{n-1}\cdot e'_n - d'e'_{n-1}\cdot e'_n) \\
&\equiv \pm\prod_{\alpha<\beta}(de_\alpha\cdot e_\beta)(d'e_\alpha\cdot e_\beta)\prod_\alpha(de_\alpha\cdot e'_{n-1} - d'e_\alpha\cdot e'_{n-1}) \\
&\qquad \cdot\wedge(de_\alpha\cdot e'_n)(de'_{n-1}\cdot e'_n) \\
&\equiv \pm\sin^{n-2}\phi\prod_{\alpha<A}\{(de_\alpha\cdot e_A)(d'e_\alpha\cdot e'_A)\}d\phi.
\end{aligned}
\tag{32}
$$

Here the congruences are to be understood mod $dP\cdot e_A$, $d'P\cdot e'_A$. The step next to the last follows from the relations

$$d'e'_A\cdot e'_n = -d'e'_n\cdot e'_A \equiv 0, \quad \bmod d'P\cdot e'_A,$$

which in turn are consequences of (18). In the reduction of the last step we make use of the relations (26), (27), and

$$de_A\cdot e_n = -de_n\cdot e_A \equiv 0, \quad \bmod dP\cdot e_A.$$

If we notice that

$$\dot{\Sigma} = \prod_i(dO'\cdot a'_i)\prod_{i<j}(da'_i\cdot a'_j),$$

and recall the expressions for $\dot{L}_V$, $\dot{L}_0$, $\dot{L}_1$, then (30) and (32) together give the formula (24).

4. Total curvature and Euler characteristic. The success of our procedure depends on the possibility of expressing the Euler-Poincaré characteristic of a domain bounded by a hypersurface Σ by an integral over Σ, a result known as the Gauss-Bonnet formula. Let Λ be the volume element of the unit hypersphere in E, and N^+ the field of outward normals of Σ. By means of N^+ we define the normal mapping of Σ. The Gauss-Bonnet formula in this particular case can be written

$$\int_{N^+}\Lambda = \chi(D)I_{n-1}, \tag{33}$$

where D is the domain bounded by Σ, and $\chi(D)$ is its Euler-Poincaré characteristic. The left-hand side of this equation is sometimes called the total curvature of the domain.

In our later application the domain D will not be bounded by a smooth

hypersurface but will be such that its boundary consists of a finite number of hypersurfaces which intersect in a number of submanifolds V^{n-2} of dimension $n-2$. To the integral of Λ over the outward normals we must then add the integral over the vectors belonging to the angle subtended by the outward normals of the two hypersurfaces. To express the latter analytically let us use the notation of the last section, together with the ranges of indices (25). In addition we denote by $\mathfrak{v}_A$, $\mathfrak{v}'_A$ the unit vectors in the principal directions of Σ_0, Σ_1 respectively. For a differentiation on Σ_0 we can then write

$$\theta_A = dP \cdot \mathfrak{v}_A, \qquad \kappa_A \theta_A = d\mathfrak{e}_n \cdot \mathfrak{v}_A, \tag{34}$$

where κ_A are the principal curvatures. Similarly, for a differentiation on Σ_1 we have

$$\theta'_A = dP \cdot \mathfrak{v}'_A, \qquad \kappa'_A \theta'_A = d\mathfrak{e}'_n \cdot \mathfrak{v}'_A, \tag{35}$$

κ'_A being the principal curvatures of Σ_1. Since the $\mathfrak{e}_\alpha$ lie in the intersection of the tangent hyperplanes, we have relations of the form

$$\mathfrak{e}_\alpha = \sum_A c_{\alpha A} \mathfrak{v}_A = \sum_A c'_{\alpha A} \mathfrak{v}'_A \tag{36}$$

To simplify notation we introduce the unit vectors $\mathfrak{v}$, $\mathfrak{w}$ in the directions of the angle bisectors of $\mathfrak{e}_n$, $\mathfrak{e}'_n$. Then we have

$$\mathfrak{e}_n = (\cos \tfrac{1}{2}\phi)\mathfrak{v} - (\sin \tfrac{1}{2}\phi)\mathfrak{w}, \qquad \mathfrak{e}'_n = (\cos \tfrac{1}{2}\phi)\mathfrak{v} + (\sin \tfrac{1}{2}\phi)\mathfrak{w}, \tag{37}$$

or

$$\mathfrak{e}_n + \mathfrak{e}'_n = 2(\cos \tfrac{1}{2}\phi)\mathfrak{v}, \qquad -\mathfrak{e}_n + \mathfrak{e}'_n = 2(\sin \tfrac{1}{2}\phi)\mathfrak{w}. \tag{38}$$

Let $\mathfrak{x}$ be a unit vector between $\mathfrak{e}_n$. $\mathfrak{e}'_n$, and $\mathfrak{y}$ the unit vector perpendicular to $\mathfrak{x}$ and in the plane of $\mathfrak{e}_n$, $\mathfrak{e}'_n$. We can then write

$$\mathfrak{x} = \cos\sigma\mathfrak{v} + \sin\sigma\mathfrak{w}, \qquad \mathfrak{y} = -\sin\sigma\mathfrak{v} + \cos\sigma\mathfrak{w}; \qquad -\tfrac{1}{2}\phi \leqq \sigma \leqq \tfrac{1}{2}\phi. \tag{39}$$

It follows that the total curvature, i. e., I_{n-1} times the Euler-Poincaré characteristic of D, is given by

$$K = \int_{N^+} \Lambda + \int_{-\frac{1}{2}\phi}^{\frac{1}{2}\phi} d\sigma \int_V \prod_\alpha \{\cos\sigma(d\mathfrak{v}\cdot\mathfrak{e}_\alpha) + \sin\sigma(d\mathfrak{w}\cdot\mathfrak{e}_\alpha)\}. \tag{40}$$

The product in the second integral admits some further simplification. In fact, using (38), we have

$$\prod_\alpha \{\cos\sigma(d\mathfrak{v}\cdot\mathfrak{e}_\alpha) + \sin\sigma(d\mathfrak{w}\cdot\mathfrak{e}_\alpha)\}$$
$$= \prod_\alpha \{\sin(\tfrac{1}{2}\phi - \sigma)(d\mathfrak{e}_n\cdot\mathfrak{e}_\alpha) + \sin(\tfrac{1}{2}\phi + \sigma)(d\mathfrak{e}'_n\cdot\mathfrak{e}_\alpha)\}/\sin^{n-2}\phi.$$

By (36), we get

$$de_n \cdot e_\alpha = \sum_A \kappa_A c_{\alpha A} \theta_A - \sum_{A,\beta} \kappa_A c_{\alpha A} c_{\beta A} (dP \cdot e_\beta),$$
$$de'_n \cdot e_\alpha = -\sum_{A,\beta} \kappa'_A c'_{\alpha A} c'_{\beta A} (dP \cdot e_\beta).$$

It follows that

$$\prod_\alpha \{\cos \sigma (d\mathfrak{v} \cdot e_\alpha) + \sin \sigma (d\mathfrak{w} \cdot e_\alpha)\} = D\dot{V}/\sin^{n-2}\phi, \tag{41}$$

where $\dot{V}$ is the volume element of V^{n-2}, and where

$$D = \left| \sin(\tfrac{1}{2}\phi - \sigma) \sum_A \kappa_A c_{\alpha A} c_{\beta A} + \sin(\tfrac{1}{2}\phi + \sigma) \sum_A \kappa'_A c'_{\alpha A} c'_{\beta A} \right|. \tag{42}$$

The determinant D can be expanded in the form

$$D = \sum_{p=0}^{n-2} H_p \sin^{n-2-p}(\tfrac{1}{2}\phi - \sigma) \sin^p(\tfrac{1}{2}\phi + \sigma), \tag{43}$$

where

$$H_p = \sum \begin{vmatrix} c_{1A_1} \cdots c_{1A_q} & c'_{1B_1} \cdots c'_{1B_p} \\ \cdots & \cdots \\ c_{n-2,A_1} \cdots c_{n-2,A_q} & c'_{n-2,B_1} \cdots c'_{n-2,B_p} \end{vmatrix}^2 \kappa_{A_1} \cdots \kappa_{A_q} \kappa'_{B_1} \cdots \kappa'_{B_p}, \quad p + q = n - 2, \tag{44}$$

the summation being extended over all independent combinations $A_1, \cdots, A_q$ and $B_1, \cdots, B_p$ of $1, \cdots, n-1$. To prove this we observe that the expansion of D is of the above form and that the question is only to determine the coefficient of $\kappa_{A_1} \cdots \kappa_{A_q} \kappa'_{B_1} \cdots \kappa'_{B_p}$ in H_p. This coefficient is, up to the factor $\sin^q(\tfrac{1}{2}\phi - \sigma) \sin^p(\tfrac{1}{2}\phi + \sigma)$, the value of D, when we set

$$\kappa_{A_1} = \cdots = \kappa_{A_q} = 1, \qquad \kappa'_{B_1} = \cdots = \kappa'_{B_p} = 1,$$

and equal to zero otherwise. Writing

$$\bar{c}_{\alpha A} = \{\sin(\tfrac{1}{2}\phi - \sigma)\}^{\frac{1}{2}} c_{\alpha A}, \qquad \bar{c}'_{\alpha A} = \{\sin(\tfrac{1}{2}\phi + \sigma)\}^{\frac{1}{2}} c'_{\alpha A},$$

we have

$$D = \left| \sum_{s=A_1,\ldots,A_q} \bar{c}_{\alpha s} \bar{c}_{\beta s} + \sum_{t=B_1,\ldots,B_p} \bar{c}'_{\alpha t} \bar{c}'_{\beta t} \right| = \begin{vmatrix} \bar{c}_{1A_1} \cdots \bar{c}_{1A_q} & \bar{c}'_{1B_1} \cdots \bar{c}'_{1B_p} \\ \cdots & \cdots \\ \bar{c}_{n-2,A_1} \cdots \bar{c}_{n-2,A_q} & \bar{c}'_{n-2,B_1} \cdots \bar{c}'_{n-2,B_p} \end{vmatrix}^2.$$

This shows that the coefficient is actually the one asserted in (43), (44).

5. Proof of the kinematic formula. Let Σ_0, Σ_1 be two hypersurfaces twice differentiably imbedded in E, with Σ_0 fixed and Σ_1 moving. We denote by D_i the domain bounded by Σ_i, $i = 0, 1$, and suppose that the intersection $D_0 \cdot D_1$ consists of a finite number of components F_s. The boundary of ΣF_s consists of the sets $\Sigma_1 \cdot D_0$, $\Sigma_0 \cdot D_1$, $\Sigma_0 \cdot \Sigma_1$, so that we can write

(45) $$\int K(D_0 \cdot D_1)\dot{\Sigma}_1 = \int K(\Sigma F_s)\dot{\Sigma}_1$$
$$= \int K(\Sigma_1 \cdot D_0)\dot{\Sigma}_1 + \int K(\Sigma_0 \cdot D_1)\dot{\Sigma}_1 = \int K(\Sigma_0 \cdot \Sigma_1)\dot{\Sigma}_1.$$

The first two integrals are easily evaluated. Take, for instance, the second integral. For every position of Σ_1 the integrand $K(\Sigma_0 \cdot D_1)$ is the integral of Λ over the outward normals to Σ_0 at points of $\Sigma_0 \cdot D_1$. This domain of integration can be decomposed in a different way by first fixing a common point of D_1 and Σ_0, rotating D_1 about *this* point, and then letting this point vary over D_1 and Σ_0 respectively. The result of this iterated integration is

(46) $$\int K(\Sigma_0 \cdot D_1)\dot{\Sigma}_1 = J_n K_0 V_1 = J_n M_{n-1}{}^{(0)} V_1.$$

Similarly, using the fact that the kinematic density is invariant under the "inversion" of a motion, we have

(47) $$\int K(\Sigma_1 \cdot D_0)\dot{\Sigma}_1 = J_n K_1 V_0 = J_n M_{n-1}{}^{(1)} V_0.$$

To evaluate the third integral in (45) we use the density formula (24), and the formulas (40)-(44) for the total curvature arising from $\Sigma_0 \cdot \Sigma_1$. We get

$$\begin{aligned}\int K(\Sigma_0 \cdot \Sigma_1)\dot{\Sigma}_1 &= \int (D/\sin^{n-2}\phi)\, d\sigma \dot{V}\dot{\Sigma}_1 = (1/J_{n-2})\int (D/\sin^{n-2}\phi)\, d\sigma \dot{L}_V \dot{\Sigma}_1 \\ &= (1/J_{n-2})\int (\sin\phi) D\, d\sigma\, d\phi\, \dot{L}_0 \dot{L}_1 \\ &= b_{n-2}\int H_0 \dot{L}_0 \dot{L}_1 + \cdots + b_0 \int H_{n-2}\dot{L}_0\dot{L}_1 \\ &= a_{n-2} M_{n-2}{}^{(0)} M_0{}^{(1)} + \cdots + a_0 M_0{}^{(0)} M_{n-2}{}^{(1)}\end{aligned}$$

where the a's and b's are numerical constants. These constants can be determined if we take Σ_0, Σ_1 to be two hyperspheres of radii 1 and h respectively. This completes the proof of the kinematic formula.

UNIVERSITY OF CHICAGO.

REFERENCES.

[1] W. Blaschke, *Integralgeometrie*, Hamburg, 1936 and 1937.

[2] S. Chern and C. T. Yen, "Sulla formula principale cinematica dello spazio ad n dimensioni," *Bollettino della Unione Matematica Italiana* (2), vol. 2 (1940), pp. 434-437.

[3] N. Steenrod, *Fibre Bundles*, Princeton, 1951.

[4] Cf., for instance, Chern, *Topics in Differential Geometry* (Mimeographed notes), Institute for Advanced Study, Princeton, 1951.

COMMUTATORS OF OPERATORS.*

By Paul R. Halmos.

If H is a (complex) Hilbert space and if P and Q are operators on H (i. e. bounded linear transformations of H into itself), the *commutator* $[P, Q]$ of P and Q is defined by

$$[P, Q] = PQ - QP.$$

The *self-commutator* $[P]$ of a single operator P is defined by

$$[P] = [P^*, P] = P^*P - PP^*.$$

My purpose in this note is to make a slight contribution to our as yet very meager knowledge of what the commutator of two operators on a Hilbert space can look like. Wintner [3] proved that if P and Q are Hermitian, then $[P, Q]$ cannot be a non-zero multiple of the identity; as Putnam [1] has pointed out, Wintner's method yields the same conclusion even without the assumption that P and Q are Hermitian. Wielandt [2] obtained (by entirely different methods) a somewhat more general result, applicable to normed algebras. Wintner then asked whether or not the negative assertion that $[P, Q]$ can never be equal to the identity can be strengthened by proving that

$$\inf\{|([P, Q]x, x)| : \| x \| = 1\} = 0.$$

Putnam showed that this is always true on a finite dimensional Hilbert space and that it remains true in the infinite dimensional case if at least one of the two operators P and Q is Hermitian, or even normal, or even semi-normal. (An operator P is *semi-normal* if P^*P and PP^* are comparable with respect to the usual partial ordering of Hermitian operators.) I propose to show that, in general, the answer to Wintner's question is no. This assertion follows easily from the fact (Theorem 2) that the real (i. e., Hermitian) part of a commutator on an infinite dimensional Hilbert space may be prescribed arbitrarily. Theorem 2, in turn, is a consequence of the assertion (Theorem 1) that every Hermitian operator on an infinite dimensional Hilbert space is the sum of two self-commutators.

* Received May 25, 1951.

For a fixed Hilbert space H, let K be the set of all sequences $x = \{x_n\}$ such that $x_n \varepsilon H$, $n = 1, 2, \cdots$, and such that $\sum_n \| x_n \|^2 < \infty$. If, for any two elements x and y of K, the inner product of x and y is defined by

$$(x, y) = \sum_n (x_n, y_n),$$

then K is a Hilbert space; K is, in fact, the direct sum of countably many copies of H. Suppose that A is a Hermitian operator on H and define an operator B on K by $(Bx)_n = Ax_n$. Define another operator U on K by writing $(Ux)_1 = 0$ and $(Ux)_n = x_{n-1}$ for $n > 1$.

Lemma 1. *If $P = BU$, then $([P]x)_1 = A^2x$ and $([P]x)_n = 0$ for $n > 1$.*

Proof. It is easy to verify that the operator B is Hermitian and that the adjoint of U is defined by $(U^*x)_n = x_{n+1}$. It follows that

$$(P^*Px)_n = (U^*B^2Ux)_n = (B^2Ux)_{n+1} = A^2(Ux)_{n+1} = A^2x_n$$

and, if $n > 1$, that

$$(PP^*x)_n = (BUU^*Bx)_n = A(UU^*Bx)_n = A(U^*Bx)_{n-1} = A(Bx)_n = A^2x_n.$$

Since $(PP^*x)_1 = A(UU^*Bx)_1 = A(0) = 0$, the proof of the lemma is complete.

It is convenient to say that a subspace H of a Hilbert space K is *large* if H contains infinitely many orthogonal copies of its orthogonal complement, or, in other words, if $\dim(H) \geqq \aleph_0 \dim(K - H)$. Thus, for example, a subspace of a separable Hilbert space is large if and only if it is infinite dimensional.

Lemma 2. *A Hermitian operator with a large null space is a self-commutator.*

Proof. Suppose first that the given Hermitian operator is positive, i. e. that it can be written in the form A^2 with a Hermitian A. Let H be the closure of the range of A. Since H is the orthogonal complement of the null space of A, there is no loss of generality in assuming that the originally given Hilbert space $\tilde{H}$ contains the direct sum K of countably many copies of H, and that, moreover, H is embedded in K so that it coincides with the set of all those sequences x in K for which $x_n = 0$ whenever $n > 1$. If an operator P is defined on K, as in Lemma 1, and extended to $\tilde{H}$ by defining it to be 0 (or, for that matter, any normal operator) on the orthogonal complement $\tilde{H} - K$, then Lemma 1 implies the desired result. If the given

operator is negative, the representation can be achievd with P^* in place of P. The case of a general Hermitian operator can be treated by putting together the results of the positive and the negative cases. It suffices to note that every Hermitian operator is the *direct* sum of a positive and a negative operator, and, in case the original operator has a large null space, then the direct summands can be selected so that they too have that property.

LEMMA 3. *Every Hermitian operator on an infinite dimensional Hilbert space leaves invariant at least one large subspace with a large orthogonal complement.*

Proof. The underlying Hilbert space, if it is not already separable, can be expressed as a direct sum of separable, infinite dimensional subspaces invariant under the given operator. There is, therefore, no loss of generality in restricting attention to separable Hilbert spaces. If A is Hermitian and E is the spectral measure of A, and if, for every Borel subset M of the real line, $E(M) = 0$ or 1, then A is a scalar multiple of 1. It follows easily that if, for every M, the dimension of the range of $E(M)$ is finite or co-finite, then A differs from a scalar multiple of 1 by a finite dimensional operator. In the contrary case both $E(M)$ and $1 - E(M)$ have infinite dimensional ranges for some M. In either case the conclusion of the lemma is obvious.

THEOREM 1. *Every Hermitian operator on an infinite dimensional Hilbert space is the sum of two self-commutators.*

Proof. By Lemma 3, the given operator is the sum of two Hermitian operators with large null spaces, and the theorem follows from Lemma 2.

To apply these results to a general operator P, it is necessary to break up P into its real and imaginary parts, i. e. the uniquely determined Hermitian operators A and B for which $P = A + iB$. If $P = A + iB$, $Q = C + iD$ (with A, B, C, and D Hermitian), it is convenient to write $P' = P'(P, Q) = A + iD$, $Q' = Q'(P, Q) = B + iC$. It follows that $P'' = P'(P', Q') = A + iC$, $Q'' = Q'(P', Q') = D + iB$, and finally that $P''' = P'(P'', Q') = P$, $Q''' = Q'(P'', Q'') = Q$. The reason for introducing P' and Q' is notational convenience; in terms of them it is easy to write down the commutator of P and Q. It is, in fact, a matter of automatic computation to verify that

$$[P, Q] = 2^{-1}([P'] + [Q']) + (2i)^{-1}([P''] + [Q'']).$$

Since a self-commutator is always Hermitian, and since an operator uniquely determines its real and imaginary parts, it follows that, for instance, the real part of $[P, Q]$ is $2^{-1}([P'] + [Q'])$. Since the transformation carrying P

and Q into P' and Q' is cyclic of order 3, it follows that any one of the three pairs $\{P, Q\}$, $\{P', Q'\}$, and $\{P'', Q''\}$ uniquely determines both others. These facts, combined with Theorem 1, yield the following result.

THEOREM 2. *Every Hermitian operator on an infinite dimensional Hilbert space is the real part of a commutator.*

COROLLARY. *There exist operators P and Q such that*

$$\inf\{|([P, Q]x, x)| : \| x \| = 1\} \geqq 1.$$

Proof. Theorem 2 yields the existence of two operators P and Q such that the real part of $[P, Q]$ is the identity. It follows that $([P, Q]x, x)$ where $\| x \| = 1$, is a complex number whose real part is 1, and that, consequently, $|([P, Q]x, x)| \geqq 1$.

It might be worth while, in closing, to call attention to another consequence of Theorem 1. Since a scalar multiple of a commutator is again a commutator, Theorem 1 and the decomposition of an operator into its real and imaginary parts imply that every operator on an infinite dimensional Hilbert space is the sum of four commutators. It follows that every additive functional of such operators, that vanishes on all commutators, vanishes identically, or, in other words, that the concept of *trace* cannot be extended to operators on infinite dimensional Hilbert spaces. (This comment was called to my attention by Irving Kaplansky.) Results of this type were known before, but only under additional assumptions of continuity or positiveness.

UNIVERSITY OF CHICAGO.

REFERENCES.

[1] C. R. Putnam, "On commutators of bounded matrices," this JOURNAL, vol. 73 (1951), pp. 127-131.

[2] H. Wielandt, "Ueber die Unbeschränktheit der Operatoren der Quantenmechanik," *Mathematische Annalen*, vol. 121 (1949), p. 21.

[3] A. Wintner, "The unboundedness of quantum-mechanical matrices," *Physical Review*, vol. 71 (1947), pp. 738-739.

ON HOMOTOPY GROUPS OF FUNCTION SPACES.*

By JAMES R. JACKSON.

1. Introduction. Let x_0 be a point of subset X_0 of topological space X, and let Y_0 be a subset of space Y. Let $y_0 \varepsilon Y_0$, and denote by the same symbol y_0 any constantly y_0-valued function. Throughout this paper, the following notations will be used:

Ω is the space of continuous mappings $f:(X, X_0) \to (Y, Y_0)$; that is, of mappings on X into Y which carry X_0 into Y_0.

Ω_0 is the space of continuous mappings $f:(X, X_0) \to (Y, y_0)$.

Ω_{00} is the space of continuous mappings $f:(X, X_0, x_0) \to (Y, Y_0, y_0)$.

Ψ is the space of continuous mappings $f: X_0 \to Y_0$.

Ψ_0 is the space of continuous mappings $f: (X_0, x_0) \to (Y_0, y_0)$.

We shall show (Section 10) that if X_0 is a retract of X, and if X, X_0, and Y satisfy certain rather general conditions (Sections 3, 4, and 5); then the m-th homotopy group $\Pi_m(\Omega, y_0)$ is isomorphic to a split extension of $\Pi_m(\Omega_0, y_0)$ by $\Pi_m(\Psi, y_0)$; and also that $\Pi_m(\Omega_{00}, y_0)$ is isomorphic to a split extension of $\Pi_m(\Omega_0, y_0)$ by $\Pi_m(\Psi_0, y_0)$.

(Group G is a *split extension of normal subgroup N by group H* if there exists a homomorphism of G onto H, with kernel N, and which induces an isomorphism of a subgroup H_0 of G onto H. It is well-known that if H_0 is also a normal subgroup, then G is the direct sum of H_0 and N.)

These results, together with some corollaries, enable us to relate the homotopy groups of many function-spaces to the homotopy groups of Y, and also to investigate certain homotopy classification problems. In particular, we provide a systematic approach to the structure of Fox's torus homotopy groups (Section 12), and list some miscellaneous interesting results (Section 13).

We also show (Section 8) that if X_0 is a deformation retract of X, and if X, X_0, and Y satisfy a weak restriction; then $\Pi_m(\Omega, y_0)$ is isomorphic to $\Pi_m(\Psi, y_0)$.

2. Some definitions. The subset $\{x \mid 0 \leq x_i \leq 1, i = 1, \cdot \cdot \cdot, m\}$ of

* Received June 6, 1951.

Euclidean m-space will be denoted by I^m. For I^1 we simply write I. The subset of point of I^m having at least one coordinate either zero or one will be designated by B^{m-1}.

Our definitions of homotopy, relative homotopy, homotopy groups, induced homomorphisms on homotopy groups, and other such concepts will be those of Fox [2] and Hu [4]. Our notation is essentially that of Hu.

Whenever a space of mappings is considered as a topological space, its topology will be the compact-open topology of Arens [1] and Fox [3].

3. Condition I. The condition discussed in the present section will be hypothesized in the main lemmas and theorems to follow. That it is not very restrictive is indicated by (3.4).

(3.1) *Definition.* We say that *X and Y satisfy Condition I* provided that whenever $\sigma: I^m \to Y^X$ is a continuous mapping, we may define a continuous mapping $\sigma^*: I^m \times X \to Y$ by

$$\sigma^*(t, x) = \sigma(t)(x) \qquad (t, x) \in I^m \times X. \tag{3.2}$$

Fox [3] has shown that for arbitrary X and Y, if $\sigma^*: I^m \times X \to Y$ is continuous, then (3.2) defines a continuous mapping $\sigma: I^m \to Y^X$. Thus, if X and Y satisfy Condition I, then (3.2) determines a one-one correspondence between the space of continuous functions on I^m into Y^X and the space of continuous functions on $I^m \times X$ into Y. Simple calculations show that this correspondence may be restricted to give a one-one correspondence between the continuous mappings $\sigma: (I^m, B^{m-1}) \to (\Omega, y_0)$ and the continuous mappings

$$\sigma^*: (I^m \times X, I^m \times X_0, B^{m-1} \times X) \to (Y, Y_0, y_0).$$

One also concludes easily from Condition I that homotopies of the functions σ relative to $\{B^{m-1}, y_0\}$ are equivalent to homotopies of the corresponding functions σ^* relative to $\{I^m \times X_0, Y_0; B^{m-1} \times X, y_0\}$.

These considerations, with parallel ones concerning Ω_0, yield the following lemma, which has been informally stated and used by Hu[7] in a more restricted case.

(3.3) LEMMA. *If X and Y satisfy Condition* I, *then the groups* $\Pi_m(\Omega, y_0)$ and $\Pi_m(\Omega_0, y_0)$, *respectively, may be considered to have as elements the homotopy classes relative to $\{I^m \times X_0, Y_0; B^{m-1} \times X, y_0\}$ of the continuous mappings*

$$\sigma^*: (I^m \times X, I^m \times X_0, B^{m-1} \times X) \to (Y, Y_0, y_0),$$

and the homotopy classes relative to $\{I^m \times X_0, y_0; B^{m-1} \times X, y_0\}$ *of the continuous mappings*

$$\sigma^*:(I^m \times X, I^m \times X_0, B^{m-1} \times X) \rightarrow (Y, y_0, y_0).$$

One sees easily that if α_1 and α_2 are each members of either of the groups mentioned in the lemma, and if they are represented by σ^*_1 and σ^*_2 according to the lemma, then the sum $\alpha_1 + \alpha_2$ is represented by $\sigma^*_3 = \sigma^*_3(t, x)$, defined by $\sigma^*_1(2t_1, t_2, \cdots, t_m, x)$ if $0 \leq t_1 \leq \frac{1}{2}$, and by $\sigma^*_2(2t_1 - 1, t_2, \cdots, t_m, x)$ if $\frac{1}{2} \leq t_1 \leq 1$.

The generality of Conditions I is indicated by the following theorem, a direct consequence of a theorem of Fox [3].

3.4) · THEOREM. *X and Y satisfy Condition* I *if either* (i) *X satisfies the first axiom of countability, or* (ii) *X is locally compact and regular (No restriction on Y in either case).*

4. Condition II. This is another hypothesis of our principal theorems. Its generality is indicated by (4.2).

Closed subset X_0 of space X is said to have the *homotopy extension property in X relative to space Y* if whenever continuous mappings $\phi: X \rightarrow Y$ and $\phi': I \times X_0 \rightarrow Y$ satisfy $\phi'(0, x) = \phi(x)$ for $x \varepsilon X_0$; then ϕ' has a continuous extension $\phi'': I \times X \rightarrow Y$ such that $\phi''(0, x) = \phi(x)$ for $x \varepsilon X$.

(4.1) *Definition.* We say that *X, X_0, and Y satisfy Condition II* provided that for $m = 1, 2, \cdots$, the subset $(I^m \times X_0) \cup (B^{m-1} \times X)$ of $I^m \times X$ has the homotopy extension property in $I^m \times X$ relative to Y.

The significance of this condition will appear in the proofs. The weakness of the restriction is indicated by the following theorem, which is a combination of [4, 9.2-9.5] with some standard theorems (For definitions of *ANR* and *ANR**, see [4]).

(4.2) THEOREM. *X, X_0, and Y satisfy Condition II if X_0 is closed in X, and if also any one of the four following requirements is met:*

(i) *X and X_0 ANR's.*

(ii) *X metric, Y an ANR.*

(iii) *X a Hausdorff space, $I^m \times X$ normal for $m = 1, 2, \cdots, Y$ an ANR*.*

(iv) *$I^m \times X$ normal for $m = 1, 2, \cdots, Y$ a compact ANR*.*

5. Conditions I and II. If X and Y, and also X_0 and Y satisfy Condition I, we say *X, X_0, and Y satisfy Condition I.* If also X, X_0. and Y satisfy Condition II, then we say that *X, X_0, and Y satisfy Conditions I and II.*

We shall mainly be interested in the case that X_0 is a retract of X (observe that this is always the case when X_0 reduces to a single point). To clear the air in this case, we set down a corollary which follows from (3.4), (4.2), and some standard theorems.

(5.1) THEOREM. *Let X_0 be a retract of X. Then X, X_0, and Y satisfy Conditions I and II if either* (i) *X is an ANR, or* (ii) *X is metric and Y is an ANR.*

These conditions are of special interest since every locally-finite polyhedron is an *ANR* [8].

6. Some homomorphisms. Define $\theta : \Omega \to \Psi$ by $\theta(\phi) = \phi | X_0$, $\phi \in \Omega$. Define $j : \Omega_0 \to \Omega$ by $j(\phi) = \phi$, $\phi \in \Omega_0$. The functions θ and j are obviously continuous.

(6.1) LEMMA. *If X, X_0, and Y satisfy Conditions I and II, then the induced homomorphism $j^* : \Pi_m(\Omega_0, y_0) \to \Pi_m(\Omega, y_0)$ carries $\Pi_m(\Omega_0, y_0)$ onto the kernel of the induced homomorphism $\theta^* : \Pi_m(\Omega, y_0) \to \Pi_m(\Psi, y_0)$.*

Proof of (6.1). It is obvious that the image of j^* is contained in the kernel of θ^*. Hence we must show that if $\alpha \in \Pi_m(\Omega, y_0)$, and if $\theta^*(\alpha) = 0$, then for some $\beta \in \Pi_m(\Omega_0, y_0)$, we have $\alpha = j^*(\beta)$. Lemma (3.3) reduces this to the following proposition.

(6.2) Let $\sigma : (I^m \times X, I^m \times X_0, B^{m-1} \times X) \to (Y, Y_0, y_0)$ be continuous, and suppose there exists a mapping $\sigma' : I \times I^m \times X_0 \to Y_0$ such that $\sigma'(0, t, x) = \sigma(t, x)$, $(t, x) \in I^m \times X_0$; $\sigma'(I \times B^{m-1} \times X_0) = y_0 = \sigma'(1 \times I^m \times X_0)$. Then there exists a mapping $\sigma'' : I \times I^m \times X \to Y$ such that

$$\sigma''(0, t, x) = \sigma(t, x),\ (t, x) \in I^m \times X; \qquad \sigma''(I \times I^m \times X_0) \subset Y_0$$
$$\sigma''(I \times B^{m-1} \times X) = y_0 = \sigma''(1 \times I^m \times X_0).$$

For suppose (6.2) is true. If σ represents an element α of the kernel of θ^*, then the hypotheses of (6.2) are fulfilled, whence σ'' exists. Define $\sigma_1 : I^m \times X \to Y$ by $\sigma_1(t, x) = \sigma''(1, t, x)$, $(t, x) \in I^m \times X$. Clearly σ_1 represents α. Since $\sigma_1(I^m \times X_0) = y_0$, the function σ_1 also represents some element $\beta \in \Pi_m(\Omega_0, y_0)$. One sees easily from the definition of j^* that $j^*(\beta) = \alpha$, as required.

Proof of (6.2). Extend σ' to $I \times [(I^m \times X_0) \cup (B^{m-1} \times X)]$ by setting $\sigma'(I \times B^{m-1} \times X) = y_0$. Obviously the extended σ' is continuous and satisfies

$$\sigma'(0, t, x) = \sigma(t, x), \qquad (t, x) \in (I^m \times X_0) \cup (B^{m-1} \times X).$$

Hence by Condition II, σ' has an extension $\sigma'': I \times I^m \times X \to Y$ such that $\sigma''(0, t, x) = \sigma(t, x)$, $(t, x) \in I^m \times X$. The function σ'' obviously satisfies the requirements of the conclusion of (6.2).

7. An important lemma. In this section we do not need Condition I or Condition II, but we shall require the more restrictive hypothesis that X_0 be a retract of X. Lemma (7.1)—in a restricted form suggested by the fact that Y may be considered as a retract of Y^X—was the starting point of the present investigation.

Let $\rho_0: X \to X_0$ be a retraction of X onto X_0. Define $\rho: \Psi \to \Omega$ by $\rho(\phi) = \phi\rho_0$, $\phi \in \Psi$. One sees easily that ρ is continuous. Let $\rho^*: \Pi_m(\Psi, y_0) \to \Pi_m(\Omega, y_0)$ be the homomorphism induced by ρ.

(7.1) LEMMA.

(i) $\theta^*\rho^*: \Pi_m(\Psi, y_0) \to \Pi_m(\Psi, y_0)$ *is the identity automorphism.*

(ii) $\rho^*: \Pi_m(\Psi, y_0) \to \Pi_m(\Omega, y_0)$ *is an isomorphism into.*

(iii) $\theta^*: \Pi_m(\Omega, y_0) \to \Pi_m(\Psi, y_0)$ *is an onto homomorphism.*

Proof of (7.1). Conclusion (i) is the fruit of a simple calculation; (ii) and (iii) are elementary set-theoretical consequences of (i).

8. An isomorphism theorem. The following result, which is not needed for the ensuing theory, has obvious generalizations relating to the concept of homotopy type.

(8.1) THEOREM. *If X, X_0, and Y satisfy Condition I, and if X_0 is a deformation retract of X; then $\theta^*: \Pi_m(\Omega, y_0) \to \Pi_m(\Psi, y_0)$ is an isomorphism onto.*

Proof of (8.1). By (7.1)(iii), we need only show that θ^* has kernel 0. Let $\tau: I \times X \to X$ retract X onto X_0 by deformation. Lemma (3.3) reduces (8.1) to the following proposition.

(8.2) Let $\sigma: (I^m \times X, I^m \times X_0, B^{m-1} \times X) \to (Y, Y_0, y_0)$, and suppose there exists a mapping $\sigma': I \times I^m \times X_0 \to Y_0$ such that

$$\sigma'(0, t, x) = \sigma(t, x), \qquad (t, x) \in I^m \times X_0;$$

$$\sigma'(I \times B^{m-1} \times X_0) = y_0 = \sigma'(1 \times I^m \times X_0).$$

Then there exists a mapping $\sigma'': I \times I^m \times X \to Y$ such that

$$\sigma''(0, t, x) = \sigma(t, x), \ (t, x) \in I^m \times X; \quad \sigma''(I \times I^m \times X_0) \subset Y_0;$$
$$\sigma''(I \times B^{m-1} \times X) = y_0 = \sigma''(1 \times I^m \times X).$$

Proof of (8.2). Define $\sigma''(s, t, x)$ to be $\sigma(t, \tau(2s, x))$ if $0 \leq s \leq \frac{1}{2}$ and $\sigma'(2s - 1, t, \tau(1, x))$ if $\frac{1}{2} \leq s \leq 1$. It is easily verified that σ'' satisfies the requirements of the conclusion of (8.2).

We set down ahead of time a simple corollary of (8.1) and (9.1).

(8.3) COROLLARY. *Let X, X_0, and Y satisfy Conditions I and II, and suppose X_0 is a deformation retract of X. Then $\Pi_m(\Omega_0, y_0)$ is trivial (that is, consists of a single element).*

9. The main lemma. The following result sharpens (6.1) for the case that X_0 is a retract of X.

(9.1) LEMMA. *If X, X_0, and Y satisfy Conditions I and II, and if X_0 is a retract of X; then the homomorphism j^* carries $\Pi_m(\Omega_0, y_0)$ isomorphically onto the kernel of θ^*.*

Proof of (9.1). By (6.1), we need only show that j^* has kernel 0. Let $\rho: X \to X_0$ be a retraction of X onto X_0. By means of (3.3), we reduce (9.1) to the following proposition.

(9.2) Let $\sigma: (I^m \times X, I^m \times X_0, B^{m-1} \times X) \to (Y, y_0, y_0)$ be continuous, and suppose there exists a mapping $\sigma': I \times I^m \times X \to Y$ such that

$$\sigma'(0, t, x) = \sigma(t, x), \ (t, x) \in I^m \times X; \quad \sigma'(I \times I^m \times X_0) \subset Y_0;$$
$$\sigma'(I \times B^{m-1} \times X) = y_0 = \sigma'(1 \times I^m \times X).$$

Then there exists a mapping $\sigma'': I \times I^m \times X \to \mathrm{Y}$ satisfying the same requirements, and also $\sigma''(I \times I^m \times X_0) = y_0$.

Proof of (9.2). Define

$$\begin{aligned}
&\tau(r, s, t, x) = y_0, && r \in I, 0 \leq s \leq r, t \in I^m, x \in X_0;\\
&\tau(r, s, t, x) = \sigma'(s - r, t, x), && r \in I, r \leq s \leq 1, t \in I^m, x \in X_0;\\
&\tau(I \times I \times B^{m-1} \times X) = y_0;\\
&\tau(r, 0, t, x) = \sigma'(0, t, x), && r \in I, t \in I^m, x \in X;\\
&\tau(r, 1, t, x) = \sigma'(1 - r, t, \rho(x)), && r \in I, t \in I^m, x \in X.
\end{aligned}$$

It is easily seen that τ is single-valued and continuous, and is defined

on $I \times [(I^{m+1} \times X_0) \cup (B^m \times X)]$. Hence by Condition II, it has an extension $\tau': I \times I^{m+1} \times X = I \times I \times I^m \times X \to Y$. Define $\sigma''(s, t; x) = \tau'(1, s, t, x)$, $(s, t, x) \varepsilon I \times I^m \times X$. One sees without difficulty that σ'' satisfies the requirements of the conclusion of (9.2).

10. The main theorems. The first theorem of this section is merely a combination of (7.1) and (9.1). The second theorem is a purely algebraic consequence of the first.

(10.1) THEOREM. *Let X, X_0, and Y satisfy Conditions I and II, and let X_0 be a retract of X. Then $\Pi_m(\Omega, y_0)$ is isomorphic to a split extension of $\Pi_m(\Omega_0, y_0)$ by $\Pi_m(\Psi, y_0)$.*

(10.2) THEOREM. *Let X, X_0, and Y satisfy Conditions I and II, and let X_0 be a retract of X. Then:*

(i) *$\Pi_m(\Omega, y_0)$ has a subgroup isomorphic to $\Pi_m(\Psi, y_0)$, and a normal subgroup isomorphic to $\Pi_m(\Omega_0, y_0)$.*

(ii) *If either of $\Pi_m(\Psi, y_0)$ and $\Pi_m(\Omega_0, y_0)$ consists of a single element, then $\Pi_m(\Omega, y_0)$ is isomorphic to the other.*

(iii) *If $\Pi_m(\Omega, y_0)$ is abelian, then it is isomorphic to the direct sum $\Pi_m(\Psi, y_0) + \Pi_m(\Omega_0, y_0)$.*

Note that the hypothesis of (10.2) (iii) is always fulfilled for $m \geqq 2$, and is fulfilled for all $m \geqq 1$ if Y is a topological group and Y_0 a subgroup [6].

Following through the proofs on which (10.1) is based, one sees easily that the following theorem can be established in exactly the same way.

(10.3) THEOREM. *Let X, X_0, and Y satisfy Conditions I and II, and let X_0 be a retract of X. Then $\Pi_m(\Omega_{00}, y_0)$ is isomorphic to a split extension of $\Pi_m(\Omega_0, y_0)$ by $\Pi_m(\Psi_0, y_0)$.*

Theorem (10.3) has, of course, a corollary parallel to (10.2).

We now provide a theorem concerning the relative homotopy groups $\Pi_m(\Omega, \Omega_0, y_0)$ and $\Pi_m(\Omega_{00}, \Omega_0, y_0)$, where Ω_0 is considered as a subset of Ω and of Ω_{00} in the obvious way. The proof, which follows easily from (7.1) (iii), (9.1), and the exactness of the homotopy sequence, is omitted. For relevant definitions and theorems, see [2] or [4, pp. 80-82, 94-99].

(10.4) THEOREM. *Let X, X_0, and Y satisfy Conditions I and II, and let X_0 be a retract of X. Then for $m = 2, 3, \cdots$, $\Pi_m(\Omega, \Omega_0, y_0)$ is isomorphic to $\Pi_m(\Psi, y_0)$, and $\Pi_m(\Omega_{00}, \Omega_0, y_0)$ is isomorphic to $\Pi_m(\Psi_0, y_0)$.*

Those familiar with relative homotopy theory and the homotopy sequence might suspect that (10.1) and (10.3) might be more easily established by first proving (10.4) directly. However, this does not seem to be the case, since existing theorems do not apply conveniently to the lemmas which arise in setting up a direct proof of (10.4).

11. Factor spaces. The theorems of Section 10 do not lend themselves to calculations of any generality because of the appearance of the group $\Pi_m(\Omega_0, y_0)$, which is difficult to deal with directly. We get around this by introducing the *factor space* $X^* = X/X_0$, which is the space obtained from X by identifying the points of X_0 to the single point x^*. The map of identification $m_0: X \to X^*$ is defined by $m_0(x) = x$, $x \in X - X_0$, and $m_0(X_0) = x^*$. The set X^* is topologized by taking subset U open if and only if $m_0^{-1}(U)$ is an open subset of X [9].

Let Ω^*_0 be the space of continuous functions $f:(X^*, x^*) \to (Y, y_0)$. It is not difficult to see that if X_0 is compact, then a homeomorphism $m: \Omega^*_0 \to \Omega_0$ is defined by setting $m(\phi) = \phi m_0$, $\phi \in \Omega^*_0$. Then the following theorem is obvious.

(11.1) THEOREM. *Let X_0 be compact. Then $\Pi_m(\Omega_0, y_0)$ is isomorphic to $\Pi_m(\Omega^*_0, y_0)$.*

We shall not enumerate the obvious corollaries to the theorems of Section 10.

12. An application. In this section we shall apply (10.1) and (11.1) to obtain the results of Fox [2] concerning the algebraic structure of his torus homotopy groups.

(12.1) DEFINITION. *Let T^0 be a point, and for $r = 1, 2, \cdots$, let T^r be the r-fold topological product of* 1*-spheres. Define $\tau_r^m(Y, y_0) = \Pi_m(Y^{T^{r-1}}, y_0)$, $m, r = 1, 2, \cdots$.*

Fox pointed out that $\tau_r^1(Y, y_0)$ *is identical with his r-th torus homotopy group of Y at base-point y_0.*

Consider T^{r-1} to be parametrized by $r - 1$ real numbers modulo 1. Define $\rho: T^{r-1} \to T^{r-1}$, $(r \geq 2)$, by $\rho(x) = (0, x_2, \cdots, x_{r-1})$, $x \in T^{r-1}$. Plainly ρ is a retraction of T^{r-1} onto a (compact) subset homeomorphic to T^{r-2}. Hence by (10.1) and (11.1), we conclude that $\tau_r^m(Y, y_0)$ is isomorphic to a split extension of $\Pi_m(\Omega^*_0, y_0)$ by $\tau^m_{r-1}(Y, y_0)$, where we see without difficulty that $\Pi_m(\Omega^*_0, y_0)$ is isomorphic to $\Pi_{m+1}(Y^{T^{r-2}}, y_0)$. This proves the following proposition.

(12. 2) *$\tau_r^m(Y, y_0)$ is isomorphic to a split extension of $\tau_{r-1}^{m+1}(Y, y_0)$ by $\tau_{r-1}^m(Y, y_0)$, for $r \geq 2$.*

Now $\tau_{r-1}^{m+1}(Y, y_0)$ is commutative, so we can apply (12. 2) to it to obtain a direct sum decomposition. Remembering that T^0 is a single point, so that $\tau_1^m(Y, y_0)$ is isomorphic to $\Pi_m(Y, y_0)$, we obtain the following direct sum decomposition by a simple induction.

$$\tau_{r-1}^{m+1}(Y, y_0) \simeq \sum_{j=2}^{r} \binom{r-2}{j-2} \Pi_{m+j-1}(Y, y_0). \tag{12. 3}$$

Fox's structure theorem is obtained by combining (12. 2) and (12.3) for the case $m = 1$.

13. Further applications. Many of the known results on homotopy groups of spaces of inessential functions are easy corollaries to the theorems of Sections 10 and 11, which also open to investigation the homotopy theories of many functions spaces inaccessible to existing results.

In view of (3. 3), it is clear that any theorem on homotopy groups of function spaces can be interpreted as a homotopy classification theorem of a special type.

We give some simple applications of our theorems. The first two are slight generalizations of known results [5], but are included for completeness, since they are needed for the other applications.

Let S^k be the k-sphere ($k \geq 0$: S^0 is a pair of points), and let s_k be a fixed point of S^k.

$$\Pi_m(Y^{S^k}\{s_k, y_0\}, y_0) \simeq \Pi_{m+k}(Y, y_0). \tag{13. 1}$$

(13. 2) *$\Pi_m(Y^{S^k}\{s_k, Y_0\}, y_0)$ is isomorphic to a split extension of $\Pi_{m+k}(Y, y_0)$ by $\Pi_m(Y_0, y_0)$.*

Proof of (13. 1). By (11. 1), we have, for $k > 0$,

$$\Pi_m(Y^{S^k}\{s_k, y_0\}, y_0) \simeq \Pi_m(Y^{I^k}\{B^{k-1}, y_0\}, y_0).$$

That

$$\Pi_m(Y^{I^k}\{B^{k-1}, y_0\}, y_0) \simeq \Pi_{m+k}(Y, y_0),$$

is an immediate consequence of (3. 3). The case $k = 0$ is trivial.

Proof of (13. 2). Take $X = S^k$ and $X_0 = s_k$ in (10. 1). (13. 2) follows

at once from (13.1) and the obvious fact that $Y_0^{s_k}$ is essentially identical with Y_0.

The following result has numerous obvious generalizations, whose proofs are similar to that below.

(13.3) *Let X be the union of S^{i-1} and S^{j-1}, joined together by identifying the points s_{i-1} and s_{j-1} to a single point x_0. Then*

$$\Pi_1(Y^X\{x_0, y_0\}, y_0) \simeq \Pi_i(Y, y_0) + \Pi_j(Y, y_0).$$

Proof of (13.3). Applying (10.3) with $X_0 = S^{i-1}$, (11.1), and (13.1), we see that $\Pi_1(Y^X\{x_0, y_0\}, y_0)$ is isomorphic to a split extension of $\Pi_i(Y, y_0)$ by $\Pi_j(Y, y_0)$. Another application of the some theorems with $X_0 = S^{j-1}$ shows that the isomorphic image of $\Pi_j(Y, y_0)$ in $\Pi_1(Y^X\{x_0, y_0\}, y_0)$ is a normal subgroup, whence the proposition follows.

We state one simple result concerning homotopy classification. Its proof follows from (13.2), when we observe that the function space with which it is concerned is homeomorphic to the space of representatives given by (3.3) for $\Pi_1(Y^{S^k}\{s_k, Y_0\}, y_0)$.

(13.4) *The homotopy classes relative to $\{s_k \times S^1, Y_0; S^k \times s_1, y_0\}$ of the space of continuous functions*

$$f: (S^k \times S^1, s_k \times S^1, S^k \times s_1) \rightarrow (Y, Y_0, y_0)$$

can be put in a 1-1 *correspondence with some split extension of* $\Pi_{k+1}(Y, y_0)$ *by* $\Pi_1(Y_0, y_0)$, and hence with the direct sum $\Pi_{k+1}(Y, y_0) + \Pi_1(Y_0, y_0)$.

The preceding applications of this section are almost obvious intuitively, although a rigorous proof is in each case except (13.1) rather difficult, without the theorems of Sections 10 and 11. The following proposition is less accessible to the imagination.

(13.5) *Let X_p be the closed orientable surface of genus p. Then* $\Pi_m(Y^{X_p}, y_0) = G_0$ *has a normal series of subgroups* $G_0 \supset G_1 \supset \cdots \supset G_{2p+1}$, *where:*

(i) *G_0 is isomorphic to a split extension of G_1 by* $\Pi_m(Y, y_0)$;

(ii) *G_i is isomorphic to a split extension of G_{i+1} by* $\Pi_{m+1}(Y, y_0)$, *for* $i = 1, \cdots, 2p$;

(iii) $G_{2p+1} \simeq \Pi_{m+2}(Y, y_0)$.

Proof of (13.5). The case $p=0$ is contained in (13.2). If $p>0$, consider X_p as a sphere with p handles. Single out a handle H, and parametrize it in the obvious way by longitude θ $(-\pi/2 \leq \theta \leq \pi/2)$ and latitude ϕ $(-\pi < \phi \leq \pi)$. Designate the point whose coordinates are θ and ϕ by (θ, ϕ).

In (10.1), take X_0 the single point $(0,0) \varepsilon H$. We find that G_0 is isomorphic to a split extension of

$$\Pi_m(Y^{X_p}\{(0,0), y_0\}, y_0) = G'_1 \tag{13.6}$$

by a group essentially identical with $\Pi_m(Y, y_0)$.

We may construct X_p in such a way that $(-\pi/2, 0)$ can be joined in X_p to $(\pi/2, 0)$ by an arc of a great circle in the surface of the sphere. Let X_0 be the union of this arc with the points of H for which $\phi=0$. Then $(0,0) \varepsilon X_0 \subset X_p$, and examination of the figure makes it clear that X_0 is a 1-sphere and is a retract of X_p. Then by (10.3) and (13.1), G'_1 is isomorphic to a split extension of $\Pi_m(Y^{X_p}\{X_0, y_0\}, y_0) = G'_2$ by $\Pi_{m+1}(Y, y_0)$.

Let X^*_p be the space obtained from X_p by identifying the points of X_0 to the single point x^*. By (11.1), G'_2 is isomorphic to $\Pi_m(Y^{X^*_p}\{x^*, y_0\}, y_0)$. Let X^*_0 be the image in X^*_p under the map of identification of the set of points in H such that $\theta=0$. One sees easily from a sketch that X^*_0 is a 1-sphere and is a retract of X^*_p. By (10.3) and (13.1), G'_2 is isomorphic to a split extension of $\Pi_m(Y^{X^*_p}\{X^*_0, y_0\}, y_0) = G'_3$ by $\Pi_{m+1}(Y, y_0)$. Using (11.1) we see easily that if $x \varepsilon X_{p-1}$, then G'_3 is isomorphic to $\Pi_m(Y^{X_{p-1}}\{x, y_0\}, y_0)$.

Thus. in essence, we are back to (13.6), except that p is reduced by unity. The proof is completed by a simple induction, concluding in an appeal to the case $p=0$, and followed by the identification of the groups G'_i with subgroups of G_0.

We note in closing that the theorems of Section 10, by reducing problems concerning spaces like Ω and Ω_{00} to problems concerning Ψ, Ψ_0, and Ω_0, greatly extend the range of application of the theorems recently obtained by Hu [7], relating homotopy groups of certain function spaces to certain subgroups of cohomology groups. Our principal theorems are, incidentally, generalizations of certain theorems of [7], but were deduced without knowledge of that paper.

UNIVERSITY OF CALIFORNIA,
LOS ANGELES.

BIBLIOGRAPHY.

[1] Arens, R. F., "A topology for spaces of transformations," *Annals of Mathematics*, vol. 47 (1946), pp. 480-495.

[2] Fox, R. H., "Homotopy groups and torus homotopy groups," *ibid.*, vol. 49 (1947), pp. 471-510.

[3] ———, "On topologies for function spaces," Bulletin of the American Mathematical Society, vol. 51 (1945), pp. 429-432.

[4] Hu, S. T., "Homotopy Theory," vol. 1, Dittoed typescript prepared by the Department of Mathematics, Tulane University, under Office of Naval Research Contract N7-ONR-434 Task Order III, 1950.

[5] ———, "On spherical mappings in a metric space," *Annals of Mathematics*, vol. 48 (1947), pp. 717-734.

[6] ———, "Some homotopy properties of topological groups and homogeneous spaces," *ibid.*, vol. 49 (1948), *pp.* 67-74.

[7] ———, "Structure of the homotopy groups of mapping spaces," *American Journal of Mathematics*, vol. 71 (1949), pp. 574-586.

[8] Liao, S. D., "On non-compact absolute neighborhood retracts," *Academia Sinica Science Record*, vol. 2 (1949), pp. 249-262.

[9] Seifert, H., and Threlfall, W., *Lehrbuch der Topologie*, Leipzig, 1934 (especially pp. 31-35).

A REMARK ON ISOLATED CRITICAL POINTS.*

By Erich H. Rothe.

1. **Introduction.** Let $I = I(x)$ be a real valued function of the point x of a space E to be specified later. We assume that the zero point o of E is an isolated critical point, i. e., that

$$\text{grad}\, I(x) = o \qquad \text{for } x = o \tag{1.1}$$

while grad $I \neq o$ for all $x \neq o$ of some neighborhood of o. In many investigations about critical points the following property which we formulate as a "hypothesis" plays a decisive role:

Hypothesis H. There exists a neighborhood U of o such that for all $x \neq o$ of the intersection [1] $U \wedge \{I(x) = I(o)\}$ the vectors $x - o$ and grad $I(x)$ are linearly independent.

If E is the (real) Euclidean n-space E^n, H is known to be true under either of the following two conditions: (i) o is a non-degenerate critical point [2]; (ii) $I(x)$ is analytic in the neighborhood of o.[3]

In a recent paper [4] hypothesis H serves as the main assumption of the theorem that for $E = E^n$ the alternating sum of the type numbers of the critical point equals the index of the singularity o of the vector field grad $I(x)$, and without proof it has been stated in this paper that the following condition is sufficient for the validity of the hypothesis H: there is an integer $p \geqq 2$ such that I has continuous differentials up to and including order $p + 2$; all differentials of order less than p vanish at $x = o$, while the homogeneous form of degree p giving the p-th differential at $x = o$ is not degenerate in the algebraic sense. This condition will be called "non-degeneracy of order p" [5] since for $p = 2$ it coincides essentially with the customary non-degeneracy condition.

* Received May 16, 1950.

[1] The symbol $\wedge$ denotes intersection, and for any property P, the symbol $\{P(x)\}$ denotes the set of all x having the property P.

[2] [6], p. 155, Theorem 4. 2.

[3] [1], Lemma 10; [6], p. 156, Theorem 4. 3.

[4] [7].

[5] Definition 2. 6.

The object of the present paper is then to prove the theorem that non-degeneracy of order p is sufficient for the validity of the hypothesis H (Theorem 3.1). The proof will be given for the case that E is a (not necessarily separable) Hilbert space which of course includes the case of a Euclidean *n-space*. The proof is given in Section 3 while Section 2 contains preliminary definitions and facts about differentials and gradients in a Hilbert space not all of which are new.

2. On differentials and gradients in the Hilbert space E. Let E be a Hilbert space, i. e., a real Banach space in which for any couple x, y of elements, a scalar product (x, y) is defined which satisfies the usual rules and is such that $(x, x)^{\frac{1}{2}}$ is the norm $\| x \|$ of the element x of E.

Definition 2.1. Let $f(x)$ be a continuous map of some open convex subset C of E into a Hilbert space E_1.[6] We define inductively differentials $d^0, d^1, d^2, \cdots$ of f as follows: $d^0 f(x) = f(x)$ for $x \varepsilon C$. Suppose now that for some integer $i \geqq 1, d^{i-1} = d^{i-1} f(x, h_1, h_2, \cdots, h_{i-1})$ has been defined for all i-tuples $(x, h_1, h_2, \cdots, h_{i-1})$ of elements of E such that $x + h_1 + h_2 + \cdots + h_{i-1} \varepsilon C$. The i-th differential d^i is then defined if and only if there exists a map $d^i = d^i f(x, h_1, \cdots, h_{i-1}, h_i)$ mapping all those $(i+1)$-tuples $(x, h_1, h_2, \cdots, h_{i-1}, h_i)$ of elements of E such that $x + h_1 + \cdots + h_{i-1} + h_i \varepsilon C$ into E_1 which has the following properties: d^i is linear[7] in h_i and

$$(2.1) \quad d^{i-1} f(x + h_i, h_1, \cdots, h_{i-1}) - d^{i-1} f(x, h_1, \cdots, h_{i-1}) = d^i f(x, h_1, \cdots, h_{i-1}, h_i) + \epsilon(x, h_1, \cdots, h_{i-1}, h_i)$$

with

$$(2.2) \qquad \lim \epsilon(x, h_1, \cdots, h_{i-1}, h_i) / \| h_i \| = 0, \text{ where } \| h_i \| \to 0.$$

If a $d^i f(x, h_1, \cdots, h_{i-1}, h_i)$ with these properties exists it is uniquely determined[8] and is called the i-th differential d^i of f in C. $d^i f(x, h_1, \cdots, h_i)$ is i-linear in $h_1, \cdots, h_i$. In addition we have the important

LEMMA 2.1. *If $d^i f(x, h_1, h_2, \cdots, h_i)$ is continuous in the argument x for all x of a neighborhood of the point x_0 then it is symmetric in $h_1, h_2, \cdots, h_i$ at $x = x_0$.*

[6] In this paper $f(x)$ will be either a real valued function (i. e., E_1 the real line) or a map of C into $E_1 = E$.

[7] "Linear" means additive *and continuous.*

[8] [3], Lemma 11.1.

For the proof see [2], Theorem 8 or [4], Satz 1.

Definition 2.2. If $Q(h_1, h_2, \cdots, h_i)$ is an i-linear symmetric real function we set $Q(h) = Q(h_1, h_2, \cdots, h_i)_{h_1=h_2 \ldots =h_i=h}$, and call $Q(h)$ a form of degree i. Correspondingly, if $f(x)$ is real-valued (i.e., E_1 the real axis) and $d^if(x, h_1, \cdots, h_i)$ exists in C we define $d^if(x, h)$ for $x + ih \varepsilon C$, by setting $d^if(x, h) = d^if(x, h_1, h_2, \cdots, h_i)_{h_1=h_2 \ldots =h_i=h}$.

If f has continuous differentials up to and including order $(n+1)$ in C then the Taylor formula

$$(2.3) \qquad f(x+h) - f(x) = \sum_{j=1}^{n} d^jf(x, h)/j! + R_{n+1};$$

$$(2.4) \qquad R_{n+1} = \int_0^1 (1-t)^n/n! d^{n+1}f(x+th, h)\, dt$$

holds if x and $x + h$ are in C.[9]

So far we have used only the Banach space property of E. We now make use of the basic property of a Hilbert space E that to every linear functional $l(x)$ there exists a uniquely determined element $g \varepsilon E$ such that $l(x) = (g, x)$ where (g, x) denotes the scalar product of the elements g and x. This property together with the linearity of d^if in h_i makes the following definition possible:

Definition 2.3. If the real-valued $f(x)$ has an i-th differential $d^if(x, h_1, \cdots, h_i)$ then the i-th gradient $g^i = g^i(x, h_1, \cdots, h_{i-1})$ is defined as the element of E which is uniquely determined by the equation

$$(2.5) \qquad d^if(x, h_1, \cdots, h_{i-1}, h_i) = (g^i(x, h_1, \cdots, h_{i-1}), h_i) = (g^i, h_i).$$

The i-linearity of d^if implies obviously the $(i-1)$-linearity of $g^i(x, h_1, \cdots, h_{i-1})$ in $h_1, \cdots, h_{i-1}$. In particular, the first gradient function $g^1(x)$ is called the gradient of $f(x)$ and we write

$$(2.6) \qquad g^1(x) = g(x) = \operatorname{grad} f(x);$$

$$(2.7) \qquad g^i(x, h_1, \cdots, h_{i-1}) = g^i(x, h) \text{ if } h_1 = \cdots = h_{i-1} = h.$$

LEMMA 2.2. *If $d^if(x, h)$ exists and is continuous in x at $x = x_0$, then there exists a neighborhood U of x_0 and a constant $C = C(x_0)$ such that for all $x \varepsilon U$*

[9] [2], Theorem 5. The integrals are "Riemann" integrals in the sense defined in [2], p. 166.

$$\| d^i f(x, h_1, \cdots, h_i) \| \leqq C \| h_1 \| \cdots \| h_{i-1} \| \| h_i \|,$$

$$\| g^i (x, h_1, \cdots, h_{i-1}) \| \leqq C \| h_1 \| \cdots \| h_{i-1} \|.$$

Proof. The first inequality follows from [4], Hilfssatz 2. The second inequality follows from the first by setting $h_i = g^i(x, h_1, \cdots, h_{i-1})$ in (2. 5).

LEMMA 2. 3. (a) *The i-th gradient $g^i(x, h_1, \cdots, h_{i-1})$ (Definition 2. 3) is symmetric in $h_1, \cdots, h_{i-1}$ if g^i is continuous in x.* (b) $d^{i-1}g(x, h) = g^i(x, h)$.

Proof. Since every permutation of $h_1, \cdots, h_{i-1}$ may be considered as a permutation of $h_1, \cdots, h_{i-1}, h_i$ (leaving h_i fixed), Lemma 2. 3(a) is an immediate consequence of (2. 5) and Lemma 2. 1.

To prove (b) we note that, because of the symmetry in $h_1, \cdots, h_{i-1}$,

$$d^i f(x, h_1, \cdots, h_{i-1}, h_i) = (g^i(x, h_1, \cdots, h_{i-2}, h_i), h_{i-1}),$$

$$d^{i-1} f(x, h_1, \cdots, h_{i-1}) = (g^{i-1}(x, h_1, \cdots, h_{i-2}), h_{i-1}).$$

Because of $d^i = dd^{i-1}$ it follows easily from the second equation that

$$d^i f(x, h_1, \cdots, h_{i-1}, h_i) = (dg^{i-1}(x, h_1, \cdots, h_{i-2}, h_i), h_{i-1})$$

and comparison with the first equation proves

$$g^i(x, h_1, \cdots, h_{i-2}, h_i) = dg^{i-1}(x, h_1, \cdots, h_{i-2}, h_i).$$

Recursive application of this formula yields (b).

Definition 2. 4. If $P(h_1, \cdots, h_{i-1})$ is an element of E which is symmetric and linear in the h_j, we set $P(h) = P(h_1, \cdots, h_{i-1})_{h_1 = \cdots = h_{i-1} = h}$ and call $P(h)$ a polynomial of degree $i - 1$ in h.[10]

For later reference we write the Taylor formula (2. 3), (2. 4) in terms of gradients (Definition 2. 3)

$$(2.8) \qquad f(x + h) - f(x) = \sum_{j=1}^{n} (g^j(x, h), h)/j! + R_{n+1}(x, h);$$

$$(2.9) \qquad R_{n+1}(x, h) = \int_0^1 (1 - t)^n / n! (g^{(n+1)}(x + th, h), h) dt.$$

It is easily seen that a form $Q(h)$ of degree i (Definition 2. 2) has a gradient. We define:

---

[10] [5], p. 63. The original definition goes back to Banach.

Definition 2.5. The form $Q(h)$ is non-degenerate if $\operatorname{grad} Q(h) \neq o$ for $h \neq o$. If there exists a constant $m > 0$ such that

$$\| \operatorname{grad} Q(h) \| \geqq m \qquad \text{for } \| h \| = 1, \tag{2.10}$$

$Q(h)$ is called strictly non-degenerate.

Remark to Definition 2.5. If E is the Euclidean n-space E^n the above definition of non-degeneracy coincides with the usual one: let $h^1, h^2, \cdots, h^n$ be the components of h in some coordinate system; then $Q(h)$ is degenerate if and only if the equations $\partial Q_i/\partial h = 0$ $(i = 1, 2, \cdots, n)$ have $h^1 = h^2 = \cdots = h^n = 0$ as the only common solution. In this case a non-degenerate form is obviously also strictly non-degenerate.

The following Lemmas 2.4 and 2.5 state some simple properties of differentials and gradients. We omit their simple proofs (cf. [3], p. 138).

LEMMA 2.4. *Let $f(h_1, h_2, \cdots, h_i)$ be a function of the i elements $h_1, \cdots, h_i$ of E. We assume that for $j = 1, 2, \cdots, i$ the differentials $d_j f(h_1, h_2, \cdots, h_i)$ of f with respect to h_j exist and are continuous in $(h_1, \cdots, h_i)$. Moreover, let $g_j(h_1, \cdots, h_i) = \operatorname{grad}_j f(h_1, \cdots, h_i)$ denote the gradient of $f(h_1, \cdots, h_i)$ considered as a function of h_j, such that for the increment η*

$$d_j f(h_1, h_2, \cdots, h_i, \eta) = (g_j(h_1, h_2, \cdots, h_i), \eta).$$

Finally, let $F(h) = f(h_1, h_2, \cdots, h_i)_{h_1 = \cdots = h_i = h}$. Then

$$dF(h, \eta) = \sum_{j=1}^{i} d_j f(h_1, h_2, \cdots, h_j, \eta)_{h_1 = h_2 = \cdots = h_i = h} \tag{2.11}$$

and

$$\operatorname{grad} F(h) = \sum_{j=1}^{i} \operatorname{grad}_j f(h_1, h_2, \cdots, h_i)_{h_1 = \cdots = h_i = h}. \tag{2.12}$$

LEMMA 2.5. *If, in addition to the assumption of Lemma 2.4, $f(h_1, h_2, \cdots, h_i)$ is symmetric in its arguments, then*

$$d_1 f(h, h, \cdots, h, \eta) = d_2 f(h, h, \cdots, h, \eta) = \cdots = d_i f(h, h, \cdots, h, \eta),$$

$$\operatorname{grad}_1 f(h, h, \cdots, h) = \operatorname{grad}_2 f(h, h, \cdots, h) = \cdots = \operatorname{grad}_i f(h, h, \cdots, h),$$

and consequently (see (2.11), (2.12)):

$$dF(h, \eta) = i\, d_i f(h, h, \cdots, h, \eta), \quad \operatorname{grad} F(h) = i \operatorname{grad}_i f(h, h, \cdots, h). \tag{2.13}$$

LEMMA 2.6. *Let $f(x)$ possess an i-th differential $d^i f(x, h_1, h_2, \cdots, h_i)$ and let $g^i(x, h_1, h_2, \cdots, h_{i-1})$ be the i-th gradient of $f(x)$ (Definition 2.3).*

Then

$$\operatorname{grad}_h d^i f(x, h, \cdots, h) = i\, g^i(x, h, \cdots, h), \tag{2.14}$$

where grad_h *means the gradient operation with respect to the variable* h.

Proof. $d^i f(x, h_1, \cdots, h_{i-1}, h_i)$ is linear in h_i and therefore equal to its own differential with respect to h_i and with the increment h_i. Consequently by the definition of grad_i as the gradient with respect to h_i we have

$$\begin{aligned} d^i f(x, h, \cdots, h_{i-1}, h_i) &= d_i d^i f(x, h_1, \cdots, h_{i-1}, h_i) \\ &= (\operatorname{grad}_i d^i f(x, h_1, \cdots, h_{i-1}, h_i), h_i). \end{aligned}$$

Comparison with (2.5) shows that

$$g^i(x, h_1, \cdots, h_{i-1}) = \operatorname{grad}_i d^i f(x, h_1, \cdots, h_{i-1}, h_i).$$

We now set $h_1 = h_2 = \cdots = h_{i-1} = h$ and apply (2.13) with $f(h_1, \cdots, h_i)$ replaced by $d^i f(x, h_1, \cdots, h_i)$ and obtain immediately (2.14).

Without proof we state the "Leibnitz rule."

LEMMA 2.7. *If the maps* $\gamma(x)$, $\delta(x)$ *of* $C \varepsilon E$ *into* E *possess* j*-th differentials then the scalat product* $f(x) = (\gamma(x), \delta(x))$ *has a* j*-th differential and*

$$d^j f(x, h) = \sum_{r=0}^{j} \binom{j}{r} (d^r \gamma(x, h), d^{j-r} \delta(x, h)).$$

Definition 2.6. Let $I = I(x)$ be a real-valued function defined in some neighborhood U of $x = o$ and p an integer $\geqq 2$. Then $I(x)$ is called non-degenerate of order p at o if the differentials of I up to and including order $p + 2$ exist and are continuous in U, if the differentials of order $1, 2, \cdots, p - 1$ are 0 at $x = o$, and if the form $d^p I(o, h)$ in h of order p is non-degenerate in the sense of Definition 2.5. If, in addition, $d^p I(o, h)$ is strictly non-degenerate then I is called strictly non-degenerate of order p.

LEMMA 2.8. *Let* $I(x)$ *possess continuous differentials in some neighborhood* U *of* o. *Then* I *is non-degenerate of order* p *at* o *if and only if*

$$g^1(o, h) = g^2(o, h) = \cdots = g^{p-1}(o, h) = o, \tag{2.16}$$

$$g^p(o, h) \neq o \text{ for } h \neq o. \tag{2.17}$$

Moreover if (2.17) *is replaced by*

$$\| g^p(o, h) \| \geqq \mu \text{ for } \| h \| = 1 \tag{2.18}$$

for some positive μ *we obtain necessary and sufficient conditions for* I *to be strictly non-degenerate at* o *of order* p.

Proof. Suppose (2.16) and (2.17) are satisfied. From (2.16) and the definition (2.5) of the gradient function it follows that

$$d^1I(o,h) = d^2I(o,h) = \cdots = d^{p-1}I(o,h) = 0. \tag{2.19}$$

Moreover from (2.17) and Lemma 2.6 we see that $\mathrm{grad}_h\, d^pI(o,h) \neq o$ for $h \neq o$, i. e. (Definition 2.5), that the form $d^pI(o,h)$ is non-degenerate. Thus (2.16) and (2.17) imply that I is non-degenerate of order p. If (2.17) is replaced by (2.18) then Lemma 2.6 shows that $\| \mathrm{grad}_h\, d^pI(o,h) \| \geqq \mu p$ and we see (Definitions 2.5 and 2.6) that I is strictly non-degenerate at $x = o$.

Conversely, suppose that I is non-degenerate of order p. Then the equations (2.19) hold identically in h, and therefore $\mathrm{grad}_h\, d^iI(o,h) = o$ for $i = 1, 2, \cdots, p-1$, which, by Lemma 2.6, implies (2.16). Moreover under our present assumption $d^pI(o,h)$ is non-degenerate, i. e., $\mathrm{grad}_h\, d^pI(o,h) \neq o$ for $h \neq o$ (Definition 2.5), which, again by Lemma 2.6, implies (2.17). In the same way this lemma shows that the strict non-degeneracy of $d^pI(o,h)$ implies the existence of a $\mu > 0$ for which (2.18) is true.

3. Proof of the hypothesis *H* in case of non-degeneracy of order *p*.

THEOREM 3.1. *Let $I(x)$ be strictly non-degenerate of order p at the origin o of the Hilbert space E (Definition 2.6).*

We assume without loss of generality that

$$I(o) = 0. \tag{3.1}$$

Then the hypothesis H (see introduction) is satisfied.

Proof. We set

$$\gamma(x) = \mathrm{grad}\{(x,x)^{p/2}\}; \tag{3.2}$$ $$\gamma(x) = px(x,x)^{p/2-1}, \tag{3.3}$$

(3.3) being implied by (3.2).

If $x = h \neq o$ is an element of E such that h and $g^1(h) = \mathrm{grad}\, I(h)$ are linearly dependent, then (3.3) shows that $\gamma(h)$ and $g^1(h)$ are also linearly dependent. Consequently in the Schwarz inequality

$$|(\gamma(h), g^1(h))| \leqq \| \gamma(h) \| \, \| g^1(h) \|$$

the equality sign will hold:

$$|(\gamma(h), g^1(h))| = \| \gamma(h) \| \, \| g^1(h) \|. \tag{3.4}$$

In order to prove Theorem 3.1 we will show that (3.4) is impossible for small enough $\| h \| \neq 0$ if

$$I(h) = 0. \tag{3.5}$$

We will indeed establish the existence of positive constants C, C' such that

$$\|\gamma(x)\| \, \| g^1(h) \| \geqq C \, \| h \|^{2p-2} \tag{3.6}$$

for small enough $\| h \| \neq 0$, while on the other hand for small enough $\| h \| \neq 0$ which in addition satisfy (3.5)

$$|(\gamma(h), g^1(h))| \leqq C' \, \| h \|^{2p-1} \tag{3.7}$$

Obviously (3.6), (3.7) are (for small enough $h \neq 0$) in contradiction with (3.4), and our theorem will be proved once the existence of constants C, C' with the above properties has been demonstrated.

We start with the proof for the existence of C. In order to estimate the left member of (3.6) we first investigate $g^1(h) = \operatorname{grad} I(h)$. To this end we apply the Taylor formula in the form (2.8), (2.9) with $n = p$ to $f(x) = I(x)$ at $x = o$ and use the equation (2.16) of Lemma 2.7 and (3.1) to obtain

$$I(h) = (g^p(o, h)/p! + R_{p+1}(o, h), \tag{3.8}$$

$$R_{p+1}(o, h) = \int_0^1 (1-t)^p/p! \, (g^{p+1}(th, h), h)\, dt. \tag{3.9}$$

To find $g^1(h) = \operatorname{grad} I(h)$ we form the differential $dI(h, \epsilon)$ of $I(h)$ with the increment ϵ. We have from (2.5) and (2.14) [11]

$$\begin{aligned} d_h[(g^p(o, h), h), \epsilon] &= d_h[d^pI(o, h), \epsilon] \\ &= (\operatorname{grad}_h d^pI(o, h), \epsilon) = p(g^p(o, h), \epsilon), \end{aligned}$$

and therefore

$$\operatorname{grad}_h (g^p(o, h), h) = p g^p(o, h). \tag{3.10}$$

Moreover if for any function $I(x, h)$ of x and h, $d_1[I(x, h), \epsilon]$ and $d_2[I(x, h), \epsilon]$ denote the differentials of $I(x, h)$ corresponding to the increment ϵ with respect to x and h respectively we see that $d_h[(g^{p+1}(th, h), h), \epsilon]$ is identical with

$$\begin{aligned} d_h[d^{p+1}I(th, h), \epsilon] &= td_1[d^{p+1}I(th, h), \epsilon] + d_2[d^{p+1}I(th, h), \epsilon] \\ &= td^{p+2}I(th, h_1, h_2, \cdots, h_{p+1}, \epsilon)_{h_1=h_2=\ldots=h_{p+1}=h} \\ &\qquad + (\operatorname{grad}_h d^{p+1}I(x, h), \epsilon)_{x=th} \\ &= t(g^{p+2}(th, h), \epsilon) + (p + 1(g^{p+1}(th, h), \epsilon), \end{aligned}$$

where again (2.14) has been used. It follows that

$$\operatorname{grad}_h (g^{p+1}(th, h), h) = tg^{p+2}(th, h) + (p+1) g^{p+1}(th, h), \tag{3.11}$$

[11] The index h on d indicates that the differential operation refers to the variable h.

and we obtain finally from (3. 8)-(3. 11):

$$(3.12) \quad \operatorname{grad} I(h) = g^p(o,h)1/(p-1)! + \int_0^1 (1-t)^p/p![tg^{p+2}(th,h) + (p+1)g^{p+1}(th,h)]dt.$$

Now from Lemma 2. 2 follows immediately the existence of a positive constant C_1 such that for small enough $\| h \|$

$$(3.13) \qquad \| g^p(o,h) \| \, 1/(p-1)! \leqq C_1 \| h \|^{p-1},$$

while

(3. 14) norm of the integral in (3. 12) $\leqq C_1 \| h \|^p$.

On the other hand, since I is strictly non-degenerate of order p at o, it follows from Lemma 2. 8 (see esp. (2. 18)) together with the $(p-1)$-linearity of $g^p(o, h_1, \cdots, h_{p-1})$ that

$$(3.15) \qquad \| g^p(o,h) \| \geqq \mu \| h \|^{p-1} \qquad (\mu > 0).$$

Obviously (3. 14), (3. 15), and (3. 12) together imply the existence of a postive constant C_2 such that for small enough $\| h \|$

$$(3.16) \qquad \| g^1(h) \| = \| \operatorname{grad} I(h) \| \geqq C_2 \| h \|^{p-1}.$$

This finally proves the validity of (3. 6) with $C = C_2 p$ since by (3. 3),

$$\| \gamma(h) \| = p \| h \|^{p-1}.$$

We turn to the proof of the existence of a $C' > 0$ such that (3. 7) holds for small enough h satisfying (3. 5). We set

$$(3.17) \qquad f(x) = (g^1(x), \gamma(x)),$$

where $\gamma(x)$ is defined in (3. 2), and apply the Taylor formula (2. 3), (2. 4) with $n = 2p - 2$ at $x = o$. Since $\gamma(o) = o$ we obtain

$$(3.18) \qquad (g^1(h), \gamma(h)) = \sum_{j=1}^{2p-2} d^j f(o,h)/j! + R_{2p-1};$$

$$(3.19) \qquad R_{2p-1} = \int_0^1 (1-t)^{2p-2}/(2p-2)! d^{2p-1} f(th,h) dt.$$

We claim first that all terms of the sum in (3. 18) are zero except for the last one, i. e., that

$$(3.20) \qquad d^j f(o,h) = 0 \qquad \text{for } j = 1, 2, \cdots, 2p-3.$$

To prove this we apply the Leibnitz rule (2.15) with $\delta(x)$ replaced by $g^1(x)$. We see from (3.3) that

$$d^1\gamma(x,h) = ph(x,x)^{(p-2)/2} + p(p-2)x(x,x)^{(p-4)/2}(x,h).$$

From this one proves easily by induction that $d^r\gamma(x,h)$ is a linear combination with constant coefficients (i. e., coefficients independent of x and h) of terms of the form

$$\text{(3.21a)} \qquad h(x,x)^{\frac{1}{2}\alpha_1}(x,h)^{\alpha_2}(h,h)^{\frac{1}{2}\alpha_3} \text{ and } x(x,x)^{\frac{1}{2}\beta_1}(x,h)^{\beta_2}(h,h)^{\frac{1}{2}\beta_3},$$

where α_i, β_i are non-negative integers satisfying

$$\text{(3.21b)} \quad 1+\alpha_2+\alpha_3 = \beta_2+\beta_3 = r, \quad 1+\beta_1+\beta_2 = \alpha_1+\alpha_2 = p-r-1.$$

It follows that $d^r\gamma(o,h) = o$ if $\alpha_1+\alpha_2 = p-r-1 > 0$, i. e., for $r = 0, 1, \cdots, p-2$, and (2.15) (with $\delta = g^1$) gives

$$\text{(3.22)} \qquad d^jf(o,h) = \begin{cases} 0 & \text{for } j = 0,1,\cdots,p-2, \\ \sum_{r=p-1}^{j} \mathcal{B}^j{}_r(d^r\gamma(o,h), d^{j-r}g^1(o,h) & \text{for } j \geqq p-1, \end{cases}$$

where the $\mathcal{B}$'s denote the binomial coefficients. We are interested in j-values $\leqq 2p-3$ (cf. 3.20); for these j-values and the r appearing in the sum of (3.22) we have $0 \leqq j-r \leqq 2p-3-(p-1) = p-2$. Therefore the right-hand member of (3.22) will be seen to be zero for $j \leqq 2p-3$, i. e., (3.20) will be proved, once it is shown that

$$\text{(3.23)} \qquad d^sg^1(o,h) = o \qquad \text{for } s = 0,1,\cdots,p-2.$$

To prove (3.23) we have only to observe that by Lemma 2.3(b), $d^sg^1(x,h) = g^{s+1}(x,h)$. But $g^{s+1}(o,h) = o$ for $s = 0,1,\cdots,p-2$ by Lemma 2.8 (equ. 2.16), which proves (3.23).

Thus (3.20) holds and (3.18) simplifies to

$$\text{(3.24)} \qquad (g^1(h),\gamma(h)) = d^{2p-2}f(o,h)/(2p-2)! + R_{2p-1}.$$

We apply (3.22) for $j = 2p-2$. Then $j-r = 2p-2-r \leqq p-2$ for $r \geqq p$. This together with (3.23) shows that (3.22) reduces to

$$\text{(3.25)} \qquad d^{2p-2}f(o,h) = \mathcal{B}^{2p-2}{}_{p-1}(d^{p-1}\gamma(o,h), d^{p-1}g^1(o,h)).$$

Now (3.21) shows that $d^{p-1}\gamma(o,h) = C'_1h(h,h)^{\frac{1}{2}\alpha_3}$, where $\alpha_3 = p-2$ and C'_1 is a constant. Consequently we obtain from (3.25)

$$d^{2p-2}f(o,h) = \mathcal{B}^{2p-2}{}_{p-1}C'_1(h,h)^{\frac{1}{2}p-1}(h, d^{p-1}g^1(o,h))$$

and by Lemma (2. 3b)

(3. 26) $$d^{2p-2}f(o,h) = \mathcal{B}^{2p-3}{}_{p-1}C'_1(h,h)^{2p-1}(g^p(o,h),h).$$

If now (3. 5) is satisfied we see from (3. 8), (3. 9), Lemma 2. 2 and the Schwarz inequality that

$$|(h, g^p(o,h))| = |\int_0^1 (1-t)^p (g^{p+1}(th,h),h)\,dt| \leqq C'_2 \| h \|^{p+1}$$

for some positive constant C'_2. Therefore, (3. 26) shows that

$$| d^{2p-2}f(o,h) | \leqq C'_3 \| h \|^{2p-1}$$

for a suitable $C'_3 > 0$. From this, (3. 24), (3. 19) and Lemma 2. 2 follows now obviously (3. 7) for some $C' > 0$.

UNIVERSITY OF MICHIGAN.

REFERENCES.

[1] A. B. Brown, "Relations between the critical points of a real analytic function of n independent variables," *American Journal of Mathematics*, vol. 52 (1930), pp. 251-270.

[2] L. M. Graves, "Riemann integration and Taylor's theorem in general analysis," *Transactions of the American Mathematical Society*, vol. 29 (1927), pp. 163-177.

[3] T. H. Hildebrandt and L. M. Graves, "Implicit functions and their differentials in general analysis," *ibid.*, vol. 29 (1927), pp. 127-153.

[4] M. Kerner, "Die Differentiale in der allgemeinen Analysis," *Annals of Mathematics*, vol. 34 (1933), pp. 546-572.

[5] S. Mazur and W. Orlicz, "Grundlegende Eigenschaften der polynomischen Operationen," *Studia Mathematica*, vol. 5 (1934), pp. 50-68.

[6] M. Morse, *The Calculus of Variations in the Large*, American Mathematical Society Colloquium Publications, vol. 18 (1934).

[7] E. H. Rothe, "A relation between the type numbers of a critical point and the index of the corresponding field of gradient vectors," *Mathematische Nachrichten*, vol. 4 (1950/51), pp. 12-27.

ON THE EMBEDDING OF HYPERBOLIC LINE ELEMENTS; A CORRECTION.*

By Philip Hartman and Aurel Wintner.

In our papers appearing in vol. 72 (1950), pp. 553-566 and vol. 73 (1951), pp. 876-884 of this *Journal*, which will be referred to as [1] and [2], respectively, we were dealing with the problem of local embedding of a binary ds^2 into a Euclidean 3-space in the three cases $K > 0$, $K < 0$, $K \equiv 0$ for the Gaussian curvature K of the ds^2. We now see that the treatment of the second of these three cases, that is, of the *hyperbolic* case $(K < 0)$, is vitiated by the proof given for the hyperbolic cases of Lemma 1 in [1]. The error is introduced at the end of the last sentence in that proof, lines 11-12 of p. 557, where it is implied that the integral representation of the solution of a hyperbolic differential equation by Riemann's function will produce a certain degree of differentiability. That such cannot be the case follows from the existence of "discontinuity waves" of any given order.

This has no bearing on the treatments of the elliptic $(K > 0)$ and parabolic $(K \equiv 0)$ cases in [1] and in [2]. Thus, the elliptic and parabolic cases of the Theorem in [1], p. 554, are not affected, nor are those statements in [2] which deal with the elliptic and parabolic cases (namely, (I) on pp. 876-877 and (iii) on p. 882) and the general theorems in [2], namely (i) on p. 879 and (ii) on p. 880.

On the other hand, the hyperbolic case can today be treated only by making use of the general theory of hyperbolic systems, which leads to comparatively high C^n-assumptions, and our method does not contribute anything to this case.

The Johns Hopkins University.

* Received November 28, 1951.

AN EXISTENCE THEOREM OF CALCULUS OF VARIATIONS FOR INTEGRALS ON PARAMETRIC SURFACES.*

By Lamberto Cesari.

In recent years the concept of integral $\mathfrak{J}(S) = \iint_S F(x,p)\,dudv$, $x = (x^1, x^2, x^3)$, $p = (p^1, p^2, p^3)$, on a continuous surface S in parametric form, $S: x = x(u,v)$, $(u,v) \varepsilon Q \equiv (0,1,0,1)$, has been studied by the author in connection with the theory for Lebesgue area [2, 3, 4]. In particular, various necessary conditions and also sufficient conditions for lower semicontinuity [3, 4] have been obtained, which extend, to the large class of all surfaces in parametric form and of finite Lebesgue area, the well-known theorems of Tonelli [26, vol. 1] for curves of finite length and of McShane [14] for surfaces under particular conditions.

In the present paper the problem of the existence of the absolute minimum of integrals $\mathfrak{J}(S)$ for surfaces in parametric form and of finite Lebesgue area is studied with the aim of advancing for such surfaces an existence theory based only upon geometrical considerations and the concept of lower semicontinuity. The following existence theorem is obtained (no. 25):

Each positive definite semiregular integral $\mathfrak{J}(S) = \iint_S F(x,p)\,dudv$ *has an absolute minimum in the class of all surfaces of finite Lebesgue area which are contained in a given closed bounded convex part A of the space E and whose boundary is a given closed Jordan curve C, provided that C is the boundary curve of at least one surface of finite Lebesgue area in A.*

This theorem, in which the condition of the boundedness of A can be removed (no. 42), extends an analogous theorem of Tonelli for integrals on a curve [26, vol. 2], as well as a theorem of McShane [17] for integrals on a surface $\mathfrak{J}(S) = \iint_S F(p)\,dudv$ in which the function F does not depend upon the point x in the space E_3. Another particular case is the problem of Plateau for which Rado, McShane, Courant, Douglas, Tonelli gave, a long time ago, final existence theorems, and for which the integral involved is the area integral $\mathfrak{J}(S) = \iint_S |\,p\,|\,dudv$.

* Received March 14, 1951.

The proof of the above existence theorem (§ 3) is based only on the recalled conditions for lower semicontinuity [3], on some results of Lebesgue area theory for surfaces, on previous observations of McShane [18] and C. B. Morrey [19], and, above all, on a detailed study of a new suitable smoothing process for polyhedral surfaces (§ 2). The author is indebted to Professor L. M. Graves for the very short and elegant proof (§ 2, n. 13) of Lemma 15, which supplants a previous somewhat longer one.

In order to point out the elementary and direct character of the procedure used in the present paper the direct application of any existence theorems for the Dirichlet problem (no. 43) has been avoided.

§ 1. Generalities on Surfaces and Integrals.

1. Let $a = (a^1, a^2, \cdots, a^n)$ be any point (vector) of the space E_n, and let $| a |$ be the norm of a; that is, $| a | = [(a^1)^2 + \cdots + (a^n)^2]^{1/2}$. Let $\{a_1, a_2\} = | a_1 - a_2 |$ be the euclidean distance between two vectors a_1, a_2; hence $\{a, 0\} = | a |$, where $0 = (0, 0, \cdots, 0)$. For any set $A \subset E_n$ let diam A be the diameter of A; that is, diam $A = \text{Sup}\, | x - y |$ for any $x, y \varepsilon A$. For any two sets $A, B \subset E_n$ let $\{A, B\}$ be the distance between A and B; that is, $\{A, B\} = \text{Inf}\, | x - y |$ for any $x \varepsilon A$, $y \varepsilon B$. For any set A let A_0, A^*, $\bar{A} = A + A^*$ be, respectively, the subset of the interior points, the boundary set, the closure set of A.

We have to deal especially with the space E_3 of the points $x = (x^1, x^2, x^3)$ and the space E_2 of the points $w = (u, v)$. Let $Q \equiv [0 \leq u \leq 1, 0 \leq v \leq 1]$ be the unit square of E_2 and let us fix the counterclockwise orientation on Q^* as the positive sense.

2. Let $x(w)$, $w \varepsilon Q$, be any single-valued vector-function continuous on Q. The equation $S: x = x(w)$, $w \varepsilon Q$, defines a continuous oriented Fréchet surface and the equation $C: x = x(w)$, $w \varepsilon Q^*$, defines a continuous oriented closed curve C. We say, in the following, simply that S is a surface and C is a curve. We call C the boundary curve $C = \partial S$ of the surface S.

Let $\| S, S' \|$, $\| C, C' \|$ be the Fréchet distance between two surfaces S, S', or two curves C, C'. We say that two surfaces S, S', or two curves C, C', are Fréchet-equivalent, and we indicate this fact by $S \sim S'$, or $C \sim C'$, if and only if $\| S, S' \| = 0$, or $\| C, C' \| = 0$. The following statements are well known.

LEMMA 1. i) $\| S, S' \| = \| S', S \|$; ii) $\| S, S' \| \geq 0$; iii) $\| S, S'' \| \leq \| S, S' \| + \| S', S'' \|$; iv) *if* $S \sim S_1$, $S' \sim S_1'$, *then* $\| S, S' \| = \| S_1, S_1' \|$. *The same for curves.*

LEMMA 2. [*See* 6] $\| \partial S, \partial S' \| \leq \| S, S' \|$. *Hence if* $S \sim S'$, *then also* $\partial S \sim \partial S'$.

If $S \sim S'$, $S: x = x(w), w \varepsilon Q$, $S': x = x'(w), w \varepsilon Q$, we say also, for simplicity's sake, that $x = x(w)$ and $x = x'(w)$ are different representations of the same Fréchet surface.

3. [For this no. see 1 and 6] Let $S: x = x(w), w \varepsilon Q$, be a given surface. For each point $w \varepsilon Q$ we call the point $x(w)$ the image of w on S. We indicate by $[S]$ the set of all points $x \varepsilon E_3$ such that x is the image of at least one point $w \varepsilon Q$. We say that $[S]$ is the set of points covered by the surface. Given a set $I \subset Q$, we indicate by $x(I)$ the set of points $x \varepsilon E_3$ such that x is the image of at least one point of I. We say that $x(I)$ is the image of I on S. We know the set $[S] \subset E_3$ is bounded, closed, connected, locally connected. For each point $x \varepsilon [S]$ let us denote by $S^{-1}(x)$, or the counterimage of x, the set of all points $w \varepsilon Q$ whose image is x. We know that $S^{-1}(x)$ is a closed subset of Q, hence its components γ are subcontinua of Q (possibly single points of Q). Let us call G the collection of all continua $\gamma \subset Q$ which are components of at least one set $S^{-1}(x) \subset Q$. The collection G has the following properties: i) each point $w \varepsilon Q$ belongs to one and only one continuum γ of G; ii) G is the collection of the maximal continua of Q on which the vector $x(w)$ is constant; iii) the collection G is upper semicontinuous on Q.

We say that a surface S is a *base surface* [in 1 an A-surface] if α) for any continuum $\gamma \varepsilon G$ the open set $E_2 - \gamma$ is connected. We say that a surface S is *non-degenerate* if β_1) for any continuum $\gamma \varepsilon G$ the open set $Q_0 - \gamma Q_0$ is connected; β_2) $\gamma \supset Q^*$ for a $\gamma \varepsilon G$ implies $\gamma = Q$, i. e., $x(w)$ is constant on Q. The properties α, β_1, β_2 are invariant for Fréchet equivalence. The following statement holds:

LEMMA 3. [1] *Any base surface* S *whose boundary curve* $C = \partial S$ *is a Jordan curve is non-degenerate.*

4. We have to deal in the following with subdivisions of Q in rectangles or in triangles. We say that a subdivision of Q in rectangles r is *regular* if the rectangles r are the following ones: $[\alpha_i \leq u \leq \alpha_{i+1}, \beta_j \leq v \leq \beta_{j+1}]$, $i = 0, 1, \cdots, m$, $j = 0, 1, \cdots, n$, where $0 = \alpha_0 < \alpha_1 < \cdots < \alpha_m < \alpha_{m+1} = 1$, $0 = \beta_0 < \beta_1 < \cdots < \beta_n < \beta_{n+1} = 1$, are any given numbers. We say that a subdivision of Q in triangles t is *regular* if each point $w \varepsilon Q$ which is a vertex of a triangle t is a vertex also of all adjacent triangles.

We say that a representation $S: x = x(w)$, $w \varepsilon Q$, of a surface S is

quasi-linear in Q, or that the vector $x(w)$ is quasi-linear in Q, if $x(w)$ is continuous on Q and there exists a regular subdivision of Q in triangles t such that $x(w)$ is linear on each t. We say that a surface S is a polyhedral surface if S possesses a quasi-linear representation. Each triangle $t \subset Q$ of the corresponding subdivision of Q has for image a triangle $T \subset E_3$ and we call the sum of the elementary areas of the triangles T the elementary area $a(S)$. Given any surface S, we term the Lebesgue area $L(S)$ of S the number $L(S) = \underline{\lim}\, a(\Sigma)$ for all possible polyhedral surfaces Σ when $\| \Sigma, S \| \to 0$.

5. We say that a representation $S: x = x(w), w \varepsilon Q$, of a surface S is a D-representation, or that $x(w)$ is a D-vector, if $x(w) = [x^1(u,v), x^2(u,v), x^3(u,v)]$, $w = (u,v) \varepsilon Q$, and i) $x^s(u,v)$, $s = 1, 2, 3$, are ACT functions in Q (i. e. absolutely continuous in the sense of Tonelli); ii) the derivatives $x_u^s = \partial x^s/\partial u$, $x_v^s = \partial x^s/\partial v$, $s = 1, 2, 3$ (which by i) exist a. e. in Q) are L^2-integrable in Q. Every quasi-linear representation is a D-representation.

For any D-vector $x(w)$ let us consider the vectors $x_u = [x_u^s, s = 1, 2, 3]$, $x_v = [x_v^s, s = 1, 2, 3]$ and the following expressions:

$$E = | x_u |^2 = \sum_{s=1}^{3} (x_u^s)^2, \qquad G = | x_v |^2 = \sum_{s=1}^{3} (x_v^s)^2, \qquad F = x_u \cdot x_v = \sum_{s=1}^{3} x_u^s x_v^s$$

$$J = J(w) = (J_1, J_2, J_3), \qquad J_s = x_u^{s+1} x_v^{s+2} - x_v^{s+1} x_u^{s+2}, \qquad s = 1, 2, 3,$$

where we mean $x^4 = x^1$, $x^5 = x^2$. We call $J = J(w)$ the Jacobian (vector) of the vector $x(w)$. Let us term the *Dirichlet integral* $D[x]$ (or D-integral) and *area integral* $I[x]$ the integrals

$$D[x] = (\tfrac{1}{2}) \iint_Q (E + G)\, dw,$$

$$I[x] = \iint_Q | J |\, dw = \iint_Q [EG - F^2]^{1/2} dw,$$

where $dw = dudv$. The following statement is well known:

LEMMA 4. [19 and 1, p, 72] *If a surface S possesses a D-representation $S: x = x(w), w \varepsilon Q$, then $L(S) = I[x] \leq D[x] < +\infty$.*

We say that a D-representation $S: x = x(w)$, $w \varepsilon Q$, of a surface S is *generalized conformal* if iii) $E = G$, $F = 0$ a. e. in Q. For every generalized conformal D-representation we have $(E + G)/2 = [EG - F^2]^{1/2}$ a. e. in Q and $I[x] = D[x]$.

LEMMA 5. [21 and 1, p. 83] *Every non-degenerate surface S with $L(S) < +\infty$ possesses a generalized conformal D-representation $S: x = x(w)$, $w \varepsilon Q$. For any representation of this kind we have*

$$L(S) = I[x] = D[x] < +\infty.$$

6. LEMMA 6. *Given a non-degenerate surface S, $L(S) < +\infty$, and any $\epsilon > 0$, there exists a polyhedral surface S_0 having a quasi-linear representation $S_0: x = x_0(w)$, $w \varepsilon Q$, such that*

$$\| \partial S, \partial S_0 \| \leq \| S, S_0 \| < \epsilon, \qquad L(S_0) = I[x_0] \leq D[x_0] < L(S) + \epsilon.$$

Proof. By Lemma 5 there exists a generalized conformal D-representation $S: x = x(w)$, $w \varepsilon Q$, of S and $L(S) = I[x] = D[x] < +\infty$. Let $x(w) = [x^s(u, v), s = 1, 2, 3]$, $w = (u, v) \varepsilon Q$, and let $x_n{}^s(u, v)$, $s = 1, 2, 3$, be the Stieltjes polynomials of order n of the continuous functions $x^s(u, v)$, $s = 1, 2, 3$ (the latter supposed defined in the square $(-1 \leq u \leq 1, -1 \leq v \leq 1)$ by symmetry and in the whole plane (u, v) by periodicity with period 2 with respect to u and v). Let S_n be the elementary surfaces $S: x = x_n(w)$, $w \varepsilon Q$, $n = 1, 2, \cdots$. We have (Lemma 4) $L(S_n) = I[x_n] \leq D[x_n] < +\infty$. Because the functions $x_n{}^s(u, v)$ are continuous in E_2, are ACT in each finite region and their derivatives $(x_n{}^s)_u$, $(x_n{}^s)_v$ are L^2-integrable in Q, then, by well known theorems [9, 27], we have $x_n(w) \rightrightarrows x(w)$ on Q and the integrals of $[EG - F^2]^{1/2}$ and of $E + G$ in Q, calculated with the polynomials $x_n{}^s(u, v)$, approach the corresponding integrals calculated with the functions $x^s(u, v)$, $s = 1, 2, 3$; that is, $I[x_n] \to I[x]$, $D[x_n] \to D[x]$ as $n \to \infty$. Hence there is an integer n such that $| x_n(w) - x(w) | < \epsilon/2$ for any $w \varepsilon Q$, and

$$L(S_n) = I[x_n] \leq D[x_n] < D[x] + \epsilon/2 = L(S) + \epsilon/2.$$

Now let us consider the surface S_n in the representation $S: x = x_n(w)$, $w \varepsilon Q$, $x_n(w) = [x_n{}^s(u, v), s = 1, 2, 3]$, where the $x_n{}^s$ are polynomials. A well known elementary procedure permits us to define a polyhedral surface S_0 (inscribed in S_n) and a quasi-linear representation $S_0: x = x_0(w)$, $w \varepsilon Q$, of S_0 such that $| x_n(w) - x_0(w) | < \epsilon/2$ for any $w \varepsilon Q$ and such that $| I[x_0] - I[x_n] | < \epsilon/2$, $| D[x_0] - D[x_n] | < \epsilon/2$. Therefore $| x(w) - x_0(w) | < \epsilon/2 + \epsilon/2 = \epsilon$ for any $w \varepsilon Q$ and (Lemma 5) $L(S_0) = I[x_0] \leq D[x_0] < D[x_n] + \epsilon/2 < L(S) + \epsilon$. Finally by Lemma 2 we have

$$\| \partial S, \partial S_0 \| \leq \| S, S_0 \| \leq \max | x(w) - x_0(w) | < \epsilon.$$

Lemma 6 is thereby proved.

7. Let $A \subset E_3$ be a closed subset of E_3; let $F(x, p)$, $x = (x^1, x^2, x^3)$, $p = (p^1, p^2, p^3)$, be any single-valued real function of the real vectors x, p such that

a) $F(x, p)$ is continuous as a function of (x, p) for any $x \in A$ and $|p| \neq 0$;

b) $F(x, tp) = tF(x, p)$ for any $x \in A$, $|p| \neq 0$, $t > 0$.

Let S be any surface such that $[S] \subset A$, $L(S) < +\infty$. Let $S: x = x(w)$, $w \in Q$, be any representation of S. Then a concept of Weierstrasse integral on the surface S, $\mathfrak{F}(S) = \iint_S F(x, p)\, dw$ $(dw = du\,dv)$, has been previously introduced [2], and $\mathfrak{F}(S)$ is invariant for Fréchet equivalence. If $F(x, p) = |p|$, then $\mathfrak{F}(S) = L(S)$. The following statements hold:

Lemma 7. *If S possesses the D-representation $S: x = x(w)$, $w \in Q$, then $\mathfrak{F}(S)$ is given by the classical Lebesgue integral*

$$\mathfrak{F}(S) = \iint_S F(x, p)\, dw = \iint_Q F[x(w), J(w)]\, dw,$$

where $J(w) = (J_1, J_2, J_3)$ is the Jacobian (vector) of the vector $x(w)$ (no. 5).

Lemma 8. *If S, S_n, $n = 1, 2, \cdots$, are surfaces of finite Lebesgue area and $[S], [S_n] \subset A$, $\| S_n, S \| \to 0$, $L(S_n) \to L(S)$, then $\mathfrak{F}(S_n) \to \mathfrak{F}(S)$.*

Lemma 7 is a particular case of a more general statement [2, p. 107]. Lemma 8 has been proved in [2, p. 101].

8. We say that $\mathfrak{F}(S)$ is a *positive definite* (semidefinite) integral if $F(x, p) > 0$ $(F(x, p) \geq 0)$ for any $x \in A$, $|p| \neq 0$. Let us now suppose that

c) the derivatives $F_s(x, p) = \partial F(x^1, x^2, x^3, p^1, p^2, p^3)/\partial p^s$, $s = 1, 2, 3$, exist and are continuous for any $x \in A$, $|p| \neq 0$.

We say that $\mathfrak{F}(S)$ is a *positive regular* (semiregular) integral if

$$\mathfrak{E}(x, p, \bar{p}) = F(x, \bar{p}) - \Sigma \bar{p}^s F_s(x, p) > 0 \ (\geq 0)$$

for any $x \in A$, $|p|, |\bar{p}| \neq 0$, $p \neq \bar{p}$. We say that $\mathfrak{F}(S)$ is a lower semicontinuous integral on a surface S_0 with respect to a family $\{S\}$ of surfaces S if, given $\epsilon > 0$, there exists a $\delta > 0$ such that for any surface $S \in \{S\}$, $\| S, S_0 \| < \delta$, we have $\mathfrak{F}(S) > \mathfrak{F}(S_0) - \epsilon$. The following statement has been previously proved:

Lemma 9. *If $F(x, p) > 0$ for any $x \in A$, $|p| \neq 0$, if S_0 is a given surface and $[S_0] \subset A$, $L(S_0) < +\infty$, and if $\mathfrak{E}(x, p, \bar{p}) \geq 0$ for any $x \in [S_0]$,*

$p \neq \bar{p}$, $|p|$, $|\bar{p}| \neq 0$, then the integral $\mathfrak{J}(S)$ is lower semicontinuous on the surface S_0 with respect to all surfaces S such that $[S] \subset A$, $L(S) < +\infty$.

This lemma implies that any positive definite semiregular integral $\mathfrak{J}(S)$ is lower semicontinuous on each surface S_0 such that $[S_0] \subset A$, $L(S_0) < +\infty$. Lemma 9, as well as this last statement, are particular cases of a previous theorem [3, p. 67].

Note. For the validity of Lemma 9 it is sufficient to suppose that condition c) holds for all $x \varepsilon [S_0]$ (not necessarily for all $x \varepsilon A$), as can be easily verified by inspecting the mentioned proof given in [3], especially pp. 61-68.

9. LEMMA 10. *To any surface S we can associate a base surface S_0 having the same boundary curve $\partial S_0 = \partial S$ and such that $[S_0] \subset [S]$, $L(S_0) \leq L(S)$. In addition, if $\mathfrak{J}(S)$ is any positive semidefinite integral and $[S] \subset A$, $L(S) < +\infty$, then we have also $\mathfrak{J}(S_0) \leq \mathfrak{J}(S)$.*

LEMMA 11. *Given a Jordan curve C of E_3, given $\epsilon > 0$, there exists a $\delta = \delta(C, \epsilon) > 0$ such that to each base surface S with $\| \partial S, C \| < \delta$ we can associate a non-degenerate surface S_0, such that $[S_0] \subset [S]$, $L(S_0) \leq L(S)$, $\| \partial S_0, C \| < \epsilon$. In addition, if $\mathfrak{J}(S)$ is any positive semidefinite integral and $[S] \subset A$, $L(S) < +\infty$, then we have also $\mathfrak{J}(S_0) \leq \mathfrak{J}(S)$.*

Lemma 10 is a result which is well known and often applied [16, 20, 7]; Lemma 11 has been recently proved [7]. Both results depend on the so-called retraction process for surfaces [7].

10. *We say that a given sequence $S_n : x = x_n(w)$, $w \varepsilon Q$, $n = 1, 2, \cdots$,* of representations of surfaces, that is, a given sequence of vectors, satisfies a *three points condition* on Q^* if Q^* contains for each n three points a_n, b_n, c_n, whose mutual distances exceed some fixed $\delta > 0$ and such that the three points $x_n(a_n)$, $x_n(b_n)$, $x_n(c_n)$ tend to three distinct limits as $n \to \infty$.

LEMMA 12. *If S_n, $n = 1, 2, \cdots$, are non-degenrate surfaces such that $L(S_n) < +\infty$ and $\| \partial S_n, C \| \to 0$, where C is a continuous closed curve not reduced to a single point, then there are generalized conformal D-representations $S_n : x = x_n(w)$, $w \varepsilon Q$, of the surfaces S_n, $n = 1, 2, \cdots$, satisfying a three points condition on Q^*.*

This Lemma is only a variant of Lemma 5 [1, 21, 35].

11. LEMMA 13. *Given any two positive numbers N, ϵ, there exists a positive number $\eta = \eta(N, \epsilon) > 0$ with the following property: To any D-vector $x(w)$, $w \varepsilon Q$, such that $D[x] < N$ we can associate a number $\delta = \delta[N, \epsilon, x(w)] > 0$, $\eta < \delta < \epsilon$, and a regular subdivision* (no. 4) *of Q in rectangles r whose dimensions are between δ and 2δ and such that the image of each side (not on Q^*) of the rectangles r on the surface $S: x = x(w)$, $w \varepsilon Q$, is a rectifiable curve whose length is $< \epsilon$.*

This Lemma has been proved recently [35, especially the first pages, pp. 317-319].

12. LEMMA 14. *Given a sequence of D-representations of surfaces $S_n: x = x_n(w)$, $w \varepsilon Q$, $n = 1, 2, \cdots$, such that $D[x_n] < N$, N constant, $n = 1, 2, \cdots$, and $\| \partial S_n, C \| \to 0$, where C is a closed Jordan curve, then the vectors $x_n(w)$, $n = 1, 2, \cdots$, are equicontinuous on Q^*.*

This Lemma has been proved in different ways [1, 21, 35].

§ 2. A Smoothing Process for Polyhedral Surfaces.

13. LEMMA 15. *Suppose $f(x)$ is L-integrable and non-negative on $0 \leq x \leq a$, and that $0 < \int_x^a f(\alpha) d\alpha \leq k f^2(x)$ almost everywhere on $0 \leq x < a$ and in particular for $x = 0$. Then $a \leq 2kf(0)$.*

Proof. The transformation $g = kf$ shows we may assume $k = 1$. Set $h(x) = \int_x^a f(\alpha) d\alpha$. Then $h(x) > 0$ for all $0 \leq x < a$ and $h'(x) = -f(x)$ a.e.; hence $0 < h(x) \leq [h'(x)]^2$ a. e., $h'(x) \leq -\sqrt{h(x)} < 0$ a. e., and and $h(x) \geq \int_x^a \sqrt{h(\alpha)} d\alpha$ on $0 \leq x \leq a$. Let $h_1(x) = \int_x^a \sqrt{h(\alpha)} d\alpha$, $h_2(x) = \int_x^a \sqrt{h_1(\alpha)} d\alpha$, $0 \leq x \leq a$. Then $h \geq h_1 \geq h_2$ on $0 \leq x \leq a$, and $h, h_1, h_2, -h_1', -h_2'$ are continuous on $0 \leq x \leq a$, and positive on $0 \leq x < a$; h_2'' also is continuous on $0 \leq x < a$, and $h_2''(x) = (1/2) \sqrt{h(x)}/h_1(x) \geq 1/2$ on $0 \leq x < a$. Hence, by the theorem of the mean, $-h_2'(x) \geq (a - x)/2$ and so $h_2(x) \geq (a - x)^2/4$, $f^2(0) \geq h(0) \geq h_2(0) \geq a^2/4$.

14. LEMMA 16. *If C is a simple closed polygonal in E_3, l the length of C, $\pi \subset E_2$ a simple polygon, $C: x = x(w)$, $w \varepsilon \pi^*$, a quasi-linear representation of C on π^*, then there is a polyhedral surface S whose boundary*

curve is $\partial S = C$ *and a quasi-linear representation* $S: x = X(w)$, $w \,\varepsilon\, \pi$, *of* S *on* π, *such that* $X(w) = x(w)$ *for any* $w \,\varepsilon\, \pi^*$, $a(S) \leq l^2/4$, *and* $[S]$ *is contained in the minimum convex body containing* C.

Proof. Let w_i, $i = 1, 2, \cdots, n$, $w_{n+1} = w_1$ be points of π^* such that $x(w)$ is linear on each arc $\tau_i = w_i w_{i+1}$, $i = 1, 2, \cdots, n$, of π^*. Let $x_i = x(w_i)$, $x_{n+1} = x_1$, $i = 1, 2, \cdots, n$, be the images of w_i on C. We can say that the x_i are vertices of C. Let S be the polyhedral surface sum of the triangles $T_i = x_1 x_i x_{i+1}$, $i = 2, 3, \cdots, n-1$. If d is the diameter of C, we have $d \leq l/2$ and

$$\begin{aligned} \text{area } T_i &\leq (1/2) \mid x_1 - x_i \mid \mid x_i - x_{i+1} \mid \leq (1/2) d \mid x_i - x_{i+1} \mid \\ &\leq (l/4) \mid x_i - x_{i+1} \mid, \qquad\qquad i = 2, 3, \cdots, n-1. \end{aligned}$$

Finally area $S \leq (l/4) \Sigma \mid x_i - x_{i-1} \mid \leq l^2/4$. Let $w' = h(w)$, $w' = (u', v')$, be a quasi-linear homeomorphism between π and the triangle

$$T \equiv [u' \geq 0, v' \geq 0, u' + v' \leq 1]$$

such that, if $w_i' = h(w_i)$ are the images of w_i on T^*, $i = 1, 2, \cdots, n$, we have $w_1' = (0, 0)$, $w_2' = (1, 0)$, $w_n' = (0, 1)$. All other points w_i', $i = 3, 4, \cdots, n-1$, are on the segment $(u' \geq 0, v' \geq 0, u' + v' = 1)$. We can also suppose that $h(w)$ is linear on each τ_i, $i = 1, 2, \cdots, n$. Let $X'(w')$ be the quasi-linear continuous vector, $w' \,\varepsilon\, T$, which is linear on each triangle $w_1' w_i' w_{i+1}'$ and there represents the triangle T_i, $i = 2, 3, \cdots, n-1$, and is such that $X'(w_i') = x_i$, $i = 1, 2, \cdots, n$. Let finally $X(w) = X'[h(w)]$, $w \,\varepsilon\, \pi$. The vector $X(w)$, $w \,\varepsilon\, \pi$, has all the required properties.

15. Let S be a polyhedral surface, $S: x = x(w)$, $w \,\varepsilon\, Q$, a quasi-linear representation of S on Q, $C = \partial S$ the boundary curve of S. Let x_0 be any point of E_3, $\rho > 0$ any real number, $B^* = B^*(\rho) \equiv [\mid x - x_0 \mid = \rho]$ the sphere of centre x_0 and radius ρ, $B = B(\rho) \equiv [\mid x - x_0 \mid < \rho]$ the interior of B^*. Let us suppose $C \subset B(r)$ for a given $r > 0$, and, for any $\rho \geq r$, let us consider the maximal connected subset $A = A(\rho)$ of Q containing Q^* and open with respect to Q, such that $x(w) \,\varepsilon\, B(\rho)$ for any $w \,\varepsilon\, A(\rho)$. Let R be the maximum distance of the points of $[S]$ from x_0. We have $R \geq r$.

LEMMA 17. *Under the above conditions* $A(\rho)$ *is an open connected set whose boundary is* $Q^* + \Sigma\pi_i^*$, *where* π_i, $i = 1, 2, \cdots, n$, *are simple disjoint Jordan regions, finite in number, and each* π_i^* *is a sum of elementary arcs. If* $\mathfrak{L}$ *is the image of* $\Sigma\pi_i^*$ *on* S, *then* $\mathfrak{L} \subset B^*(\rho)$ *is a sum of a finite number*

of elementary curves and, if $l(\rho)$ is the total length of $\mathfrak{L}$, we have $l(\rho-0) = l(\rho) \leq l(\rho+0)$ for any $r \leq \rho \leq R$.

Proof. The vector $x(w)$ is quasi-linear on Q, hence is linear on the triangles of a regular subdivision of Q in triangles. Let v, d, t denote the vertices, sides, triangles of the subdivision. Let V, D, T denote the images of the v, d, t. Let us consider the distances of all elements V, D, T from x_0 (in the sense of set theory, no. 1). We have a finite collection $\{\rho\}$ of numbers ρ. All elements V are points of E_3, all elements D are segments or points, all elements T are triangles, or segments, or points. For each triangle T the intersection TB^* is one (or two, or three) arcs λ of circumference (or a single point). Hence $\mathfrak{L}$, as a part of the intersection SB^*, is a sum of a finite number of closed curves p_i, $i = 1, 2, \cdots, n$, each a sum of arcs of circumference λ. This holds for any p, $r \leq \rho \leq R$. Let us denote by $\mathfrak{A} = \mathfrak{A}(\rho)$ the image of $A = A(\rho)$. Each arc λ is the image of an arc of a conic in a triangle t in Q and the curves π_i^* are sums of these arcs, plus, possibly, segments, finite in number, whose images are single points which are intersections of segments T with B^*. If ρ is not in the finite collection $\{\rho\}$, no point V is on B^*, hence for each arc λ the vertices of the triangle T to which λ belongs are not on B^*. Hence λ separates two annular strips in T, one in $\mathfrak{A}$ and in B, one not in $\mathfrak{A}$ and outside of $B + B^*$. The curves ρ_i and their total length $l(\rho)$ vary continuously as ρ varies in a sufficiently small neighborhood of the given ρ, i. e. $l(\rho-0) = l(\rho) = l(\rho+0)$.

If ρ is a number in the collection $\{\rho\}$, then suppose first that there is a vertex on B^*. If V is such a vertex, for each triangle T having a vertex in V let us consider the part of T contained in a circle of center V and sufficiently small radius. This part may be completely contained in B, or completely outside of $B + B^*$, or partially in B and partially outside of $B + B^*$; and in this last case such a part contains an arc of circumference λ ending in V which separates points in B from points outside of $B + B^*$. Therefore the neighborhood u of V on S can be divided in sectors τ alternatively in B and outside of $B + B^*$. Not all sectors τ which are in B belong to $\mathfrak{A}$. If such a sector τ is in B and belongs to $\mathfrak{A}$, then the two corresponding arcs λ ending in V are adjacent arcs on a curve p_i. If a sector τ is in B but does not belong to $\mathfrak{A}$, then τ is a part of a region Σ in S which is completely inside of B and which belongs to $\mathfrak{A}(\rho')$ for all $\rho' > \rho$. As ρ varies continuously in a left neighborhood of the given value ρ, all arcs λ vary continuously, remaining inside of $\mathfrak{A}$; hence $l(\rho-0) = l(\rho)$. As ρ varies continuously in a right neighborhood of the given value ρ, again all arcs λ

vary continuously, but we have new arcs λ, which approach the arcs λ of the boundaries of the regions Σ. This assures that $l(\rho+0)$ exists and $l(\rho+0) \geq l(\rho)$.

Analogous reasoning holds for each of the values $\rho \varepsilon \{\rho\}$ which equal the distances of segments D, or triangles T, from x_0.

We have proved that $l(\rho-0) = l(\rho) = l(\rho+0)$ for any $r \leq \rho \leq R$, ρ not $\varepsilon \{\rho\}$; $l(\rho-0) = l(\rho) \leq l(\rho+0)$ for any $r \leq \rho \leq R$, $\rho \varepsilon \{\rho\}$. Lemma 17 is thereby proved.

16. With the same notation as in no. 15 let $\sigma(\rho)$ be the collection of the elementary surfaces defined by the vector $x(w)$ on the Jordan regions π_i, $i=1,2,\cdots,n$. Let $\alpha(\rho)$ be the total area of $\sigma(\rho)$. Evidently $\alpha(R)=0$, $\alpha(\rho) \leq a(S)$, $r \leq \rho \leq R$, and $\alpha(\rho)$ is a monotone non-increasing function of ρ, $r \leq \rho \leq R$. In addition, the previous discussion (proof of Lemma 17) assures us that $\alpha(\rho)$ is continuous at each point ρ, $r \leq \rho \leq R$, ρ not $\varepsilon \{\rho\}$.

LEMMA 18. *Under the above conditions we have, for any* ρ,

$$\alpha(\rho) \geq \int_\rho^R l(\rho)\,d\rho, \qquad r \leq \rho \leq R.$$

Note. This Lemma has a quite evident and elementary proof which we are going to give in the following lines (see also 32). For a higher formulation of this Lemma, not necessary for our purpose, see [12, 13].

Proof of Lemma 18. First of all let us observe that $a(\mathfrak{A}) + a(\sigma) = a(S)$. Now let us consider any two values of ρ, $\rho_1 < \rho_2$, such that the interval (ρ_1, ρ_2) is free of points of $\{\rho\}$; then the difference $\alpha(\rho_1) - \alpha(\rho_2)$ is the sum of a finite number of the elementary areas described by the arcs of circumference $\lambda = \lambda(\rho)$ on the corresponding triangles T of S. If we consider one of these T and we call h the distance of x_0 from the plane of T, x_1 the projection of x_0 on such a plane, then $\lambda = \lambda(\rho)$ is an arc of circumference of center x_1 and radius $[\rho^2 - h^2]^{1/2}$. If we indicate with $\lambda(\rho)$ also the length of $\lambda(\rho)$ [the endpoints of $\lambda(\rho)$ are on the boundary of T], then, as ρ varies between ρ_1 and ρ_2, $\lambda(\rho)$ describes the elementary area

$$\tau = \int_{\rho_1}^{\rho_2} \lambda(\rho)\,d[\rho^2-h^2]^{1/2} = \int_{\rho_1}^{\rho_2} \lambda(\rho)\{\rho/[\rho^2-h^2]^{1/2}\}\,d\rho \geq \int_{\rho_1}^{\rho_2} \lambda(\rho)\,d\rho.$$

By adding these relations for all triangles T containing an arc λ we have

$$\alpha(\rho_1) - \alpha(\rho_2) \geq \int_{\rho_1}^{\rho_2} l(\rho)\,d\rho. \tag{1}$$

Let us suppose that ρ, $r \leq \rho \leq R$, be any given number and that $\rho \leq \rho_1 < \rho_2 < \cdots < \rho_m \leq R$ are all the distinct members of $\{\rho\}$ between ρ and R. By (1), which holds on each interval interior to (ρ_i, ρ_{i+1}), we have

$$\alpha(\rho_i + 0) - \alpha(\rho_{i+1} - 0) \geq \int_{\rho_i}^{\rho_{i+1}} l(\rho)\, d\rho, \qquad i = 1, 2, \cdots, m-1,$$

and, $\alpha(\rho)$ being monotone non-increasing, also

$$\alpha(\rho_i) - \alpha(\rho_{i+1}) \geq \int_{\rho_i}^{\rho_{i+1}} l(\rho)\, d\rho, \qquad i = 1, 2, \cdots, m-1.$$

Analogously

$$\alpha(\rho) - \alpha(\rho_1) \geq \int_{\rho}^{\rho_1} l(\rho)\, d\rho, \qquad \alpha(\rho_m) - \alpha(R) \geq \int_{\rho_m}^{R} l(\rho)\, d\rho,$$

if $\rho < \rho_1$, or $\rho_m < R$. In any case $\alpha(R) = 0$, hence, by adding,

$$\alpha(\rho) \geq \int_{\rho}^{R} l(\rho)\, d\rho.$$

17. We will prove, in nos. 19-24, the following

LEMMA 19. *Let S be a polyhedral surface, $S: x = x(w), w \varepsilon Q$, a quasi-linear representation of S on Q, $C = \partial S$ the boundary curve of S, L the length of C, D any number $\geq$ the diameter of C, x_0 any point of C, K any positive constant. Then there exists a polyhedral surface S_0 and a quasi-linear representation $S_0: x = x_0(w), w \varepsilon Q$, of S_0 on Q, with the following properties:*

i) *S_0 is contained in a sphere of center x_0 and radius*

$$\rho \leq 2D + 3[Ka(S)]^{1/2}.$$

ii) *there is an open set $\pi \subset Q$, which is the sum $\Sigma \pi_i$ of a finite number of disjoint simple polygons π_i, $i = 1, 2, \cdots, n$, such that $x_0(w) = x(w)$ on $Q - \pi$; hence $x_0(w) = x(w)$ on each π_i^*, $i = 1, 2, \cdots, n$, and on Q^*. Denote by σ, σ_0 $[\sigma_i, \sigma_{0i}]$ the polyhedral surfaces represented by $x(w)$, $x_0(w)$ on π $[\pi_i, i = 1, 2, \cdots, n]$ and by $a(\sigma)$, $a(\sigma_0)$ their total area, $\sigma = \Sigma \sigma_i$, $\sigma_0 = \Sigma \sigma_{0i}$, $a(\sigma) = \Sigma a(\sigma_i)$, $a(\sigma_0) = \Sigma a(\sigma_{0i})$. Each surface σ_{0i} is contained in the minimum convex body containing the curve $\partial \sigma_i = \partial \sigma_{0i}$. Denote by p_i the curves $\partial \sigma_i = \partial \sigma_{0i}$, and also their lengths, $i = 1, 2, \cdots, n$; denote by $\mathfrak{L}$ the family of curves p_i, $i = 1, 2, \cdots, n$, and also their total length, $\mathfrak{L} = \Sigma p_i$; then*

iii) $\mathfrak{L}^2 = (\Sigma p_i)^2 \leq K^{-1} a(\sigma)$;

iv) $a(\sigma_0) \leq (K^{-1}/4) a(\sigma)$;

v) *each continuum $c \subset Q$ such that $cQ^* \neq 0$, $\operatorname{diam} x(c) < D$, is completely contained in $Q - \pi$.*

18. *Note.* 1) If we suppose D exactly the diameter of C, then $D \leq L/2$ and the inequality i) becomes:

$$\rho \leq L + 3[Ka(S)]^{1/2}.$$

2) In the statements of Lemma 19 it is not excluded that π is empty, hence $S = S_0$, $\mathfrak{L} = 0$, $a(\sigma_0) = 0$.

19. The curve C is completely interior to any sphere $B^*(r)$ of center x_0 and radius $r > D$. We can take for r the value $r = 2D$. Let us call R the maximum distance of the points x of $[S]$ from x_0. Then S is contained in the sphere $B^*(R)$ of center x_0 and radius R. If $R \leq 2D + 3[Ka(S)]^{1/2}$ the Lemma is proved. Let us suppose therefore $R > 2D + 3[Ka(S)]^{1/2}$, hence $R > r$. Let us consider for any ρ, $r \leq \rho \leq R$, the set $A = A(\rho)$ studied in nos. 15, 16, the functions $l(\rho) \geq 0$, $\alpha(\rho) \geq 0$ $r \leq \rho \leq R$, and the finite family $\sigma = \sigma(\rho)$ of elementary surfaces contained in S and represented by the vector $x(w)$ on the finite family $\pi = \Sigma\pi_i$ of simple Jordan regions, $\pi = Q - (A + A^*)$ (no. 15). We know (Lemma 17) that $l(\rho - 0) = l(\rho) \leq l(\rho + 0)$ for any $r \leq \rho \leq R$ and that $l(\rho)$ is the total length of the boundary of σ and $\alpha(\rho)$ the total area of σ. In addition (Lemma 18), for any $r \leq \rho \leq R$, we have

$$\int_{\rho}^{R} l(\alpha)\, d\alpha \leq \alpha(\rho).$$

Let I be the set of all ρ such that $l(\rho) = 0$. The set I is not empty because $l(R) = 0$. If ρ_0 is a limit point at the left for I then $l(\rho_0 - 0) = l(\rho_0) = 0$; if ρ_0 is a limit point at the right for I, then $0 \leq l(\rho_0 - 0) = l(\rho_0) \leq l(\rho_0 + 0) = 0$, hence $l(\rho_0) = 0$. This proves that I is closed. Let us indicate by R_0, $r \leq \rho \leq R$, the smallest $\rho \in I$, $r \leq \rho \leq R$. Let us suppose $R_0 \leq 2D + 3[Ka(S)]^{1/2}$. We have $l(R_0) = 0$, i. e. the vector $x(w)$ represents on π^* curves reduced to a single point, hence $x(w)$ is constant on each simple closed curve π_i^* bounding a component π_i of π. Let us define the continuous vector $x_0(w)$ on Q by saying that $x_0(w)$ coincides with $x(w)$ in $Q - \pi$, and $x_0(w)$ is constant on each component π_i of π. We have $a(\sigma_0) = 0$, $\mathfrak{L} = 0$, and the surface S_0 is contained in the sphere $B^*(R_0)$. Because the vector $x(w)$ is quasi-linear in Q, the simple closed curves π_i^* bounding the components of π are sums of a finite number of segments, that is the components of π are simple polygons in Q; and this assures also that the vector $x_0(w)$ is quasi-linear on Q. Lemma 19 is therefore proved if $R_0 \leq 2D + 3[Ka(S)]^{1/2}$.

20. Let us suppose now $R_0 > 2D + 3[Ka(S)]^{1/2}$, hence $r < R_0 \leq R$. We have $l(\rho) > 0$ for any $r \leq \rho < R_0$. Let us observe that the interval $(r, r + [Ka(S)]^{1/2})$ is completely contained in the interval (r, R_0). Let us suppose, if possible, that $l(\rho) > [K^{-1}a(S)]^{1/2}$ for any $r \leq \rho \leq r + [Ka(S)]^{1/2}$. Then by Lemma 18 we would have

$$a(S) > \alpha(r) \geq \int_r^R l(\rho)\,d\rho > \int_r^{r+[Ka(S)]^{\frac{1}{2}}} [K^{-1}a(S)]^{1/2} d\rho = a(S),$$

which is contradictory. Hence there must be at least one number r', $r \leq r' \leq r + [Ka(S)]^{1/2}$, such that $l(r') \leq [K^{-1}a(S)]^{1/2}$. (For the reasoning in this number see [32].)

21. Let us consider the functions of ρ, $\phi(\rho) = \int_\rho^{R_0} l(\alpha)\,d\alpha$, $Kl^2(\rho)$, $r' \leqq \rho \leqq R_0$, where $\phi(\rho)$ is continuous and $l(\rho - 0) = l(\rho) \leq l(\rho + 0)$ for any ρ. Let J be the set of all numbers ρ, $r' \leq \rho \leq R_0$, such that $\phi(\rho) > Kl^2(\rho)$ and let r'' be the g. l. b. of the numbers $\rho \,\varepsilon\, J$. Let $r'' = R_0$ if J is empty.

Let us observe that for $\rho = R_0$ we have $\phi(R_0) = 0$, $l(R_0) = 0$; hence $\rho = R_0$ does not belong to J. In consequence we have $r'' = R_0$ only if J is empty.

A) If $r'' = r'$ then either $r' \,\varepsilon\, J$ and $\phi(r') > Kl^2(r')$, or there are $\rho \,\varepsilon\, J$ as close as we want to r' for which $\phi(\rho) > Kl^2(\rho)$, hence $\phi(r') \geq Kl^2(r' + 0) \geq Kl^2(r')$. In both cases

$$\int_{r''}^{R_0} l(\alpha)\,d\alpha \geq Kl^2(r''), \qquad r'' = r'.$$

B) If $r' < r'' < R_0$, then we have $\phi(\rho) \leq Kl^2(\rho)$ for any $r' \leq \rho < r''$; hence $\phi(r'') \leq Kl^2(r'' - 0) = Kl^2(r'')$. On the other hand there are numbers $\rho \,\varepsilon\, J$, $r'' < \rho < R_0$, as close as we want to r'' for which $\phi(\rho) > Kl^2(\rho)$; hence $\phi(r'') \geq Kl^2(r'' + 0)$. It follows that $l(r'' - 0) \geq l(r'' + 0)$, while we know that $l(r'' - 0) = l(r'') \leq l(r'' + 0)$. Therefore $l(r'' - 0) = l(r'') = l(r'' + 0)$ and $\phi(r'') = Kl^2(r'')$. We conclude that

$$\int_\rho^{r''} l(\alpha)\,d\alpha \leq \int_\rho^{R_0} l(\alpha)\,d\alpha \leq Kl^2(\rho)$$

for any $r' \leq \rho \leq r''$, where $l(\rho) > 0$ for any $r' \leq \rho \leq r''$ and

$$\int_{r''}^{R_0} l(\alpha)\,d\alpha = Kl^2(r'').$$

By Lemma 15 it follows that

$$r'' - r' \leq 2Kl(r') \leq 2K[K^{-1}a(S)]^{1/2} = 2[Ka(S)]^{1/2}.$$

C) If $r'' = R_0$ then J is empty, $\phi(R_0) = 0$, $l(R_0) = 0$, $\int_\rho^{R_0} l(\alpha)\,d\alpha \leq Kl^2(\rho)$ for any $r' \leq \rho \leq R_0$, $l(\rho) > 0$ for any $r' \leq \rho < R_0$, and finally $0 = \int_{R_0}^{R_0} l(\alpha)\,d\alpha = Kl^2(R_0)$. Again, by Lemma 15, we have

$$r'' - r' = R_0 - r' \leq 2Kl(r') \leq 2[Ka(S)]^{1/2}.$$

Let us observe that in all three cases we have

$$r'' \leq r' + 2[Ka(S)]^{1/2} \leq r + 3[Ka(S)]^{1/2} = 2D + 3[Ka(S)]^{1/2},$$

$$\int_{r''}^{R_0} l(\alpha)\,d\alpha \geq Kl^2(r'').$$

22. From no. 15 we know that the boundary $\mathfrak{L}$ of the surface $\sigma = \sigma(r'')$ is the sum of a finite number of closed elementary curves $p_1, p_2, \cdots, p_n$ and that each p_j is a sum of a finite number of arcs of circumference $\lambda = xx'$ of the sphere $B^*(r'')$. Each λ is contained in a triangle T of S and has its end point x, x' on the boundary T^* of T. Hence the plane region between the arc $\lambda = xx'$ and the segment $\lambda' = (xx')$ (both with endpoints x, x' on T^*) is completely contained in T, is interior to the sphere $B^*(r'')$ and belongs to the part $\mathfrak{A}(\rho)$ of the surface S. If we substitute each arc λ of the curve p_j with the segments λ' we obtain a polygonal line p_j' inscribed in p_j and completely contained in the sphere $B^*(r'')$. For each triangle T (containing an arc λ, let us consider the corresponding triangle t of Q; then λ and λ' have counterimages in t which are respectively an arc τ of a conic and a segment τ' having the same endpoints on the boundary t^* of t, and the plane region between τ and τ' belongs to t and to $A = A(r'')$, whereas τ belongs to $A^* = A^*(r'')$. Then p_j' is the image under the vector $x(w)$ of the polygonal line $\pi_j'^*$ which we obtain by substituting the arcs τ of A^* with the segments τ', $p_j' : x = x(w)$, $w \,\varepsilon\, \pi_j'^*$. This holds for any $j = 1, 2, \cdots, n$; hence we have, on Q, n simple polygons π_j', $j = 1, 2, \cdots, n$, without common interior points and, if π' is their sum, $Q - \pi' \subset A = A(r'')$, $\pi' = \Sigma\pi_j'$, $\pi_j' \supset \pi_j'$, $j = 1, \cdots, n$. The polygonal closed curves p_j', $j = 1, 2, \cdots, n$, are all contained in the sphere $B^*(r'')$ and we indicate their sum by $\mathfrak{L}'$. If $\mathfrak{L}'$ is also the total length of $\mathfrak{L}'$, we have $\mathfrak{L}' \leq l(r'')$, where $l(r'')$ is the total length of $\mathfrak{L}$. Let us call p_j' also the length of p_j', $j = 1, 2, \cdots, n$; hence $p_1' + p_2' + \cdots + p_n' = \mathfrak{L}'$. Let us denote by σ_j' the surfaces defined by the vector $x(w)$ on π_j and by σ' their sum. We have $a(\sigma') \geq a(\sigma)$, where $\sigma = \sigma(r'')$ and $a(\sigma) = \alpha(r'')$; hence $\alpha(r'') \leq a(\sigma')$.

23. For each polygonal closed curve p_j' there is (Lemma 16) a polyhedral surface σ_{0j} having p_j' as a boundary curve, and a quasi-linear representation $\sigma_{0j}: x = x_j(w)$, $w \varepsilon \bar{\pi}_j' = \pi_j' + \pi_j'^*$, such that $x_j(w) = x(w)$ for any $w \varepsilon \pi_j'^*$ and $a(\sigma_{0j}) \leq (1/4)p_j'^2$, $j = 1, 2, \cdots, n$. Let us denote by σ_0 the finite family of surfaces σ_{0j}, $j = 1, 2, \cdots, n$, and by $x_0(w)$, $w \varepsilon Q$, the continuous vector defined by $x_0(w) = x_j(w)$ for any $w \varepsilon \pi_j'$, $j = 1, 2, \cdots, n$, $x_0(w) = x(w)$ for any $w \varepsilon Q - \pi'$. The vector $x_0(w)$ is quasi-linear on Q, hence the surface $S_0: x = x_0(w)$, $w \varepsilon Q$, is a polyhedral surface. The polygonal closed curves p_j', $j = 1, 2, \cdots, n$, are contained in the sphere $B^*(r'')$ (the vertices are points on the sphere $B^*(r'')$). By Lemma 16 each surface σ_{0j} is contained in the minimum convex body containing p_j', hence each surface σ_{0j}, and consequently also the whole surface S_0, is contained in the sphere $B^*(r'')$. Thus i) is proved.

24. By Lemma 16 and all the preceding considerations, we have

$$\begin{aligned} a(\sigma_0) &= \Sigma a(\sigma_{0j}) \leq (1/4)\Sigma p_j'^2 \leq (1/4)(\Sigma p_j')^2 = (1/4)\mathfrak{L}'^2 \\ &\leq (1/4)l^2(r'') \leq (K^{-1}/4)\int_{r''}^{R_0} l(\alpha)\,d\alpha \leq (K^{-1}/4)\int_{r''}^{R_0} l(\alpha)\,d\alpha \\ &\leq (K^{-1}/4)\alpha(r'') \leq (K^{-1}/4)a(\sigma'); \end{aligned}$$

that is, $a(\sigma_0) \leq (K^{-1}/4)a(\sigma')$, and iv) is proved. We have

$$\begin{aligned} \mathfrak{L}'^2 &= (\Sigma p_j')^2 \leq l^2(r'') \leq K^{-1}\int_{r''}^{R_0} l(\alpha)\,d\alpha \leq K^{-1}\int_{r''}^{R} l(\alpha)\,d\alpha \\ &\leq K^{-1}\alpha(r'') \leq K^{-1}a(\sigma'); \end{aligned}$$

thus iii) is proved. Finally if, for a continuum $c \subset Q$, we have $cQ^* \neq 0$, $\operatorname{diam} x(c) < D$, then for any w, $\bar{w}$, $w \varepsilon c$, $\bar{w} \varepsilon cQ^*$, we have

$$\begin{aligned} | x(w) - x_0 | &\leq | x(w) - x(\bar{w}) | + | x(\bar{w}) - x(w_0) | \\ &\leq \operatorname{diam} x(c) + \operatorname{diam} C < D + D = 2D; \end{aligned}$$

hence $x(c) \subset B(2D)$ and $c \subset Q - \pi$. Thus also v) is proved. Lemma 19 is now completely proved.

§ 3. Existence Theorems.

25. THEOREM I. *Let A be a closed bounded convex set of E_3 and C a closed Jordan curve in A. Let W be the family of all oriented Fréchet surfaces S in A whose Lebesgue area is finite and whose boundary curve is C,*

and let us suppose W not empty. Then every positive definite and semi-regular integral $\mathfrak{J}(S)$ has an absolute minimum in W.

The proof follows in nos. 26-41.

26. The function $F(x,p)$ is continuous in the closed bounded set I of all points (x,p) such that $x \in A$, $|p| = 1$, and $F > 0$ for any $(x,p) \in I$. Then if m, M are the minimum and the maximum of F in I, we have $0 < m \leq F(x,p) \leq M$ for any $(x,p) \in I$ and therefore $m|p| \leq F(x,p) \leq M|p|$ for any $x \in A$ and any $|p| \neq 0$. By a theorem of McShane [18] we can define for any $x \in E_3$ and $|p| = 1$ a function $F_0(x,p)$ continuous at each point (x,p) and coinciding with $F(x,p)$ for any $(x,p) \in I$. In addition we can suppose $m \leq F_0(x,p) \leq M$ for any $x \in E_3$, $|p| = 1$. Let us define $F_0(x,p)$ for each $|p| \neq 0$ by putting $F_0(x,p) = |p| F_0(x, p/|p|)$. We have $F_0(x,p) = F(x,p)$ for any $x \in A$, $|p| \neq 0$. Indeed, by no. 7, b, $F_0(x,p) = |p| F_0(x,p/|p|) = |p| F(x,p/|p|) = F(x,p)$. The function $F_0(x,p)$ satisfies all the conditions of no. 7. By $\mathfrak{J}(S)$ we mean now $\mathfrak{J}(S) = \int\int_S F_0(x,p)\,dw$ and, for any surface S such that $[S] \subset A$, we have $\mathfrak{J}(S) = \int\int_S F(x,p)\,dw$.

We observe that $m \leq F_0(x,p) \leq M$ for any $x \in E_3$, $|p| = 1$, and $m|p| \leq F_0(x,p) \leq M|p|$ for any $x \in E_3$, $|p| \neq 0$. In consequence, by [2], we have also

$$mL(S) \leq \mathfrak{J}(S) \leq ML(S) \tag{1}$$

for any surface S.

27. For any $\delta \geq 0$ let A_δ be the closed convex bounded set of all points $x \in E_3$ such that $\{x, A\} \leq \delta$. We have $A_\delta \supset A$ and $A_0 = A$. For any $\delta \geq 0$ let $W(\delta)$ be the family of all surfaces S such that $[S] \subset A_\delta$, $\| \partial S, C \| \leq \delta$. We have $W(\delta) \supset W$, $W(0) = W$, and $W(\delta') \supset W(\delta)$ for any $0 \leq \delta \leq \delta'$. W is not empty, hence also $W(\delta)$ is not empty for any $\delta > 0$. Let us denote by $i, j(\delta)$ the g.l.b. of $\mathfrak{J}(S)$ when $S \in W$, or $S \in W(\delta)$, $\delta > 0$. Then $0 \leq j(\delta) \leq i < +\infty$ and $j(\delta') \leq j(\delta)$ for any $0 < \delta \leq \delta'$. Hence there exists $j = \lim j(\delta)$ as $\delta \to 0+0$ and we have $0 \leq j(\delta) \leq j \leq i$ for any $\delta > 0$. For any integer $n = 1, 2, \cdots$, let

$$\epsilon_n = 2^{-n}, \qquad \mu_n = \min[(n+2)^{-1}2^{-n-2}, 2^{-2}M\epsilon_n^2],$$

and let $\delta_n > 0$ be a real number such that $j(5\delta_n) > j - \mu_n$, $\delta_n \leq 1/n$.

Finally let $\alpha_n' = \alpha(C, \delta_n)$ be the number of Lemma 11 related to the Jordan curve C and the number $\epsilon = \delta_n$. Let $\alpha_n = \min[\alpha_n', \delta_n]$.

28. Let us consider the family $W(\alpha_n)$. There is a surface $S_n' \varepsilon W(\alpha_n)$ such that

$$j(\alpha_n) \leq \mathfrak{J}(S_n') \leq j(\alpha_n) + \mu_n,$$

and we have $[S_n'] \subset A_{\alpha_n}$, $\| \partial S_n', C \| \leq \alpha_n$. By (1) we have $L(S_n') \leq (1/m)\mathfrak{J}(S_n')$. By Lemma 10 the surface S_n' can be retracted to a base surface S_n'' having the same boundary $\partial S_n'' = \partial S_n'$ and such that $[S_n''] \subset [S_n']$, $L(S_n'') \leq L(S_n')$, $\mathfrak{J}(S_n'') \leq \mathfrak{J}(S_n')$. Hence $\| \partial S_n'', C \| \doteq \| \partial S_n', C \| \leq \alpha_n$. By Lemma 11 the surface S_n'' can be retracted to a non-degenerate surface S_n''' whose boundary $\partial S_n'''$ is a continuous curve such that $\| \partial S_n''', C \| \leq \delta_n$. In addition

$$[S_n'''] \subset [S_n''], \qquad L(S_n''') \leq L(S_n''), \qquad \mathfrak{J}(S_n''') \leq \mathfrak{J}(S_n'').$$

We observe that $\delta_n \leq 1/n$, hence $\| \partial S_n''', C \| \to 0$ as $n \to \infty$. On the other hand the surfaces S_n''', $n = 1, 2, \cdots$, are non-degenerate surfaces whose areas $L(S_n''')$ satisfy the relation

$$\begin{aligned} L(S_n''') \leq L(S_n'') \leq L(S_n') &\leq (1/m)\mathfrak{J}(S_n') \\ &\leq (1/m)[j(\alpha_n) + \mu_n] \leq (1/m)(j+1) < +\infty. \end{aligned}$$

By Lemma 12 there are generalized conformal D-representations S_n''': $x = x_n'''(w)$, $w \varepsilon Q$, of each S_n''' on Q, satisfying a three points condition on Q^* and (no. 5) $D[x_n'''] = L(S_n''')$, $n = 1, 2, \cdots$. By Lemma 6 and Lemma 8 there is, for each n, a polyhedral surface S_n and a quasi-linear representation $S_n : x = x_n(w)$, $w \varepsilon Q$, of S_n on Q, such that $| x_n(w) - x_n'''(w) | \leq \delta_n$ for any $w \varepsilon Q$, $D[x_n(w)] < D[x_n'''(w)] + \mu_n$, $| \mathfrak{J}(S_n) - \mathfrak{J}(S_n''') | < \mu_n$.

In the first place we have $| x_n(w) - x_n'''(w) | \leq \delta_n$ for any $w \varepsilon Q^*$; hence also the vectors $x_n(w)$, $n = 1, 2, \cdots$, satisfy a three points condition on Q^*. In addition, if $C_n = \partial S_n$ is the boundary curve of S_n, we have $\| C_n, \partial S_n''' \| = \| \partial S_n, \partial S_n''' \| \leq \delta_n$ and $\| C_n, C \| \leq \| C_n, \partial S_n''' \| + \| \partial S_n''', C \| \leq \delta_n + \delta_n = 2\delta$. Finally we have

$$[S_n] \subset [S_n''']_{\delta_n} \subset [S_n'']_{\delta_n} \subset [S_n']_{\delta_n} \subset [A_{\alpha_n}]_{\delta_n} \subset A_{\alpha_n + \delta_n} \subset A_{2\delta_n}.$$

This implies that $S_n \varepsilon W(2\delta_n)$, hence $\mathfrak{J}(S_n) \geq j(2\delta_n) \geq j - \mu_n$. By all the preceding considerations we have

$$\begin{aligned} \mathfrak{J}(S_n) \leq \mathfrak{J}(S_n''') + \mu_n \leq \mathfrak{J}(S_n'') + \mu_n &\leq \mathfrak{J}(S_n') + \mu_n \\ &\leq j(\alpha_n) + 2\mu_n \leq j + 2\mu_n, \end{aligned}$$

and, by (1) and no. 5, also

$$a(S_n) \leq D[x_n(w)] \leq D[x_n'''(w)] + \mu_n = L(S_n''') + \mu_n$$
$$< (1/m)(j+1) + 1.$$

We can assemble the properties of the polyhedral surfaces S_n and of their quasi-linear representations $S_n : x = x_n(w)$, $w \in Q$, (satisfying a three points condition on Q^*) as follows:

$$\| C_n, C \| \leq 2\delta_n, \qquad [S_n] \subset A_{2\delta_n}, \qquad S_n \in W(2\delta_n),$$
$$a(S_n) \leq D[x_n] < (1/m)(j+1) + 1, \qquad j - \mu_n \leq \mathfrak{F}(S_n) \leq j + 2\mu_n,$$
$$n = 1, 2, \cdots.$$

29. The representations $x = x_n(w)$, $w \in Q$, of S_n on Q are quasi-linear, hence they are D-representations and, by no. 28, $D[x_n(w)] < (1/m)(j+1) + 1$. The constant $N = (1/m)(j+1) + 1$ does not depend upon n. C is a Jordan curve and $\| C_n, C \| \to 0$ as $n \to \infty$, hence, by Lemma 14, the vectors $x_n(w)$, $w \in Q$, are equicontinuous on Q^*. There is a subsequence $x_{n_m}(w)$, $w \in Q$, of vectors $x_n(w)$ which converges uniformly on Q^*. Let us suppose, for simplicity's sake, that $n_m = m$. Let $X(w) = \lim x_n(w)$, $w \in Q^*$, as $n \to \infty$. If $C_0 : x = X(w)$, $w \in Q^*$, we have $\| C_n, C_0 \| \to 0$, $\| C_n, C \| \to 0$, $\| C, C_0 \| \leq \| C, C_n \| + \| C_n, C_0 \|$, hence $\| C, C_0 \| = 0$. That is, $x = X(w)$, $w \in Q^*$, gives a representation of the Jordan curve C. By the equicontinuity of the vectors $x_n(w)$, $w \in Q^*$, $n = 1, 2, \cdots$, we can determine, for each integer ν ($\nu = 1, 2, \cdots$), a real number $\tau_\nu > 0$ such that

$$| x_n(w) - x_n(w') | \leq \min[\epsilon_\nu, \delta_\nu], \qquad n = 1, 2, \cdots,$$

for any $w, w' \in Q^*$, $| w - w' | \leq \tau_\nu$. The numbers τ_ν depend upon ν but not on n.

30. Let us consider again the D-vectors $x_n(w)$, $w \in Q$, $n = 1, 2, \cdots$, which satisfy the relation $D[x_n] < N$, N constant, $n = 1, 2, \cdots$, and let us apply Lemma 13. Let ν indicate any integer, $\nu = 1, 2, \cdots$, and $\xi_\nu > 0$ be any real number. Then, by Lemma 13, with $\epsilon = \xi_\nu$, we can determine, for each ν, real numbers $\eta_\nu > 0$, $\delta_{n\nu} > 0$, $\eta_\nu < \delta_{n\nu} < \xi_\nu$, and regular subdivisions (no. 4) $\Delta_{n\nu}$ of Q in rectangles r whose dimensions lie between $\delta_{n\nu}$ and $2\delta_{n\nu}$, such that the image of each side of the rectangles $r \in \Delta_{n\nu}$ *not* on Q^*, is a rectifiable curve of length $\leq \xi_\nu$. We observe that the numbers η_ν depend only on ν, and $\delta_{n\nu}$ on ν and n.

As the application of Lemma 13 can be made successively for $\nu = 1, 2, \cdots$, we can choose

$$\xi_\nu = \min[\tau_\nu/2, \delta_\nu, \epsilon_\nu, \eta_{\nu-1}/2^3, \eta_{\nu-2}/2^4, \cdots, \eta_1/2^{\nu+1}], \qquad \nu = 1, 2, \cdots .$$

For each ν the dimensions of all rectangles $r \varepsilon \Delta_{n\nu}$, $n = 1, 2, \cdots$, lie between $\delta_{n\nu}$ and $2\delta_{n\nu}$, hence between η_ν and $2\xi_\nu < \eta_{\nu-1}$. Therefore they are less than the dimensions of all rectangles $r \varepsilon \Delta_{n,\nu-1}$, $n = 1, 2, \cdots$. The images of the sides on Q^* of the rectangles $r \varepsilon \Delta_{n\nu}$ under the vector $x_n(w)$ are arcs of diameter $\leq \min[\epsilon_\nu, \delta_\nu]$. The images of all other sides of the rectangles $r \varepsilon \Delta_{n\nu}$ under the vector $x_n(w)$ are arcs of length $\leq \min[\epsilon_\nu, \delta_\nu]$. The image of the boundary r^* of each $r \varepsilon \Delta_{n\nu}$ is a curve of diameter $\leq \min[4\epsilon_\nu, 4\delta_\nu]$ if $r^*Q^* \neq 0$, and is a curve of length $\leq \min[4\epsilon_\nu, 4\delta_\nu]$ if $r^*Q^* = 0$. In any case we have $\operatorname{diam} x_n(r^*) \leq 4\epsilon_\nu$ for any $r \varepsilon \Delta_{n\nu}$, $\nu = 1, 2, \cdots$, $n = 1, 2, \cdots$.

31. In the following we suppose n an arbitrary integer $(n = 1, 2, \cdots)$, and $\nu = 1, 2, \cdots, n$. Let us indicate (no. 4) by $w_{ij} \equiv (u_i, v_j)$, $i = 0, 1, 2, \cdots, \alpha + 1$, $j = 0, 1, 2, \cdots, \beta + 1$, $u_0 = v_0 = 0$, $u_{\alpha+1} = v_{\beta+1} = 1$, all vertices of the rectangles $r \varepsilon \Delta_{n\nu}$. Let us denote by τ each segment $(u = u_i, v_j \leq v \leq v_{j+1})$, $i = 1, \alpha$, $j = 1, 2, \cdots, \beta - 1$, or $(u_i \leq u \leq u_{i+1}, v = v_j)$, $i = 1, 2, \cdots, \alpha - 1$, $j = 1, \beta$. Let τ' be all analogous segments $\tau' \subset Q^*$ with $i = 0, \alpha + 1$, $j = 0, 1, \cdots, \beta$, or $i = 0, 1, \cdots, \alpha$ $j = 0, \beta + 1$. The segments τ constitute the boundary $R_{n\nu}^*$ of the rectangle $R_{n\nu} \equiv [u_1 \leq u \leq u_\alpha, v_1 \leq v \leq v_\beta]$, the segments τ' constitute the boundary Q^* of Q. Let us indicate by t and t' the polygonal lines which are images of the segments τ and τ' under the vector $x_n(w)$. Let $S_{n\nu}'$ be the polyhedral surface $S_{n\nu}' : x = x_n(w)$, $w \varepsilon R_{n\nu}$. Then the sums of the arcs t, or t', constitute the boundary curves $\partial S_{n\nu}'$, $\partial S_n = C$ of the surfaces $S_{n\nu}'$, S_n. Clearly $\mathfrak{J}(S_{n\nu}') \leq \mathfrak{J}(S_n)$. Let us make correspond to each segment $\tau \subset R_{n\nu}^*$ with $i = 1, \alpha$, $j = 1, 2, \cdots, \beta - 1$, the segment $\tau' \subset Q^*$ with $i = 0, \alpha + 1$ and the same j; let us make correspond to each segment $\tau \subset R_{n\nu}^*$ with $i = 1, 2, \cdots, \alpha - 1$, $j = 1, \beta$, the segment $\tau' \subset Q^*$ with $j = 0$, $\beta + 1$ and the same i (the correspondences being linear); let us make correspond to the vertices $w_{11}, w_{\alpha 1}, w_{\alpha\beta}, w_{1\beta}$ of $R_{n\nu}^*$ the arcs of Q^* which are sums of the two segments $\tau' \subset Q^*$ adjacent to the vertices $w_{00}, w_{\alpha+1,0}, w_{\alpha+1,\beta+1}, w_{0,\beta+1}$ of Q^*. We have a continuous monotone correspondence between $R_{n\nu}^*$ and Q^*. If $w \varepsilon R_{n\nu}^*$, $w' \varepsilon Q^*$ are corresponding points, we have

$$| x_n(w) - x_n(w') | \leq | x_n(w) - x_n(\bar{w}) | + | x_n(\bar{w}) - x_n(\bar{w}') | + | x_n(\bar{w}') - x_n(w') |,$$

where, if $w \varepsilon \tau$, $w' \varepsilon \tau'$, then $\bar{w}$, $\bar{w}'$ are corresponding end points of τ and τ';

analogously if w is a vertex of $R_{n\nu}^*$ and $w' \varepsilon r'$, $\bar{w} = w$. By nos. 29, 30, we have $|x_n(w) - x_n(w')| \leq \xi_\nu + \xi_\nu + \delta_\nu \leq 3\delta_\nu$. This assures us that $\| \partial S_{n\nu}', \partial S_n \| \leq 3\delta_\nu$ and therefore $\| \partial S_{n\nu}', C \| \leq \| \partial S_{n\nu}', \partial S_n \| + \| \partial S_n, C \| \leq 3\delta_\nu + 2\delta_\nu = 5\delta_\nu$. On the other hand $[S_{n\nu}'] \subset [S_n] \subset A_{2\delta_n} \subset A_{2\delta_\nu}$, and finally, by no. 27, $S_{n\nu}' \varepsilon W(5\delta_N)$ and

$$j - \mu_\nu \leq \mathfrak{I}(S_{n\nu}') \leq \mathfrak{I}(S_n) \leq j + 2\mu_n \leq j + 2\mu_\nu.$$

In addition,

$$0 \leq \mathfrak{I}(S_n) - \mathfrak{I}(S_{n\nu}') \leq 3\mu_\nu.$$

32. For each rectangle $r \varepsilon \Delta_{n\nu}$ let us call Σ the polyhedral surface $\Sigma : x = x_n(w)$, $w \varepsilon r$, and $L = \partial\Sigma$. Let L denote also the length of $\partial\Sigma$. Let us prove that $a(\Sigma) \leq (5M/m)\epsilon_\nu^2$. First let us suppose $r \subset R_{n\nu}$; thus $L \leq 4\epsilon_\nu$. If the above statement were false then, for a certain $r \varepsilon \Delta_{n\nu}$, we would have $a(\Sigma) > (5M/m)\epsilon_\nu^2$. Thus, by Lemma 16 applied to the curve $L : x = x_n(w)$, $w \varepsilon r^*$, there is a polyhedral surface Σ^* contained in the minimum convex body containing L, of area $a(\Sigma^*) \leq (\tfrac{1}{4})L^2 \leq (\tfrac{1}{4})(4\epsilon_\nu)^2 = 4\epsilon_\nu^2$ and a quasi-linear representation $\Sigma^* : x = x^*(w)$, $w \varepsilon r$, of Σ^* on r with $x^*(w) = x_n(w)$, $w \varepsilon r^*$. If we denote by $S_{n\nu}^* : x = x_{n\nu}^*(w)$, $w \varepsilon R_{n\nu}$, the surfaces that we obtain by putting $x_{n\nu}^*(w) = x_n(w)$, $w \varepsilon R_{n\nu} - r$, $x_{n\nu}^*(w) = x^*(w)$, $w \varepsilon r$, then we have $\partial S_{n\nu}^* = \partial S_n$ and $[S_{n\nu}^*]$ is contained in the same convex body $A_{2\delta_n}$ in which $S_{n\nu}$ is already contained. Hence $\| \partial S_{n\nu}^*, C \| = \| \partial S_{n\nu}', C \| \leq 5\delta_\nu$, $[S_{n\nu}^*] \subset A_{2\delta_\nu}$ and consequently $S_{n\nu}^* \subset W(5\delta_\nu)$ and $\mathfrak{I}(S_{n\nu}^*) \geq j - \mu_\nu$. Now, by no. 27, we have

$$\begin{aligned} j - \mu_\nu \leq \mathfrak{I}(S_{n\nu}^*) &= \mathfrak{I}(S_{n\nu}^*) + \mathfrak{I}(\Sigma^*) - \mathfrak{I}(\Sigma) \leq (j + 2\mu_\nu) + Ma(\Sigma^*) - ma(\Sigma) \\ &< j + 2\mu_\nu + 4M\epsilon_\nu^2 - m(5M/m)\epsilon_\nu^2 = j + 2\mu_\nu - M\epsilon_\nu^2 \\ &\leq j + 2\mu_\nu - 4\mu_\nu = j - 2\mu_\nu, \end{aligned}$$

which is contradictory. This proves $a(\Sigma) \leq (5M/m)\epsilon_\nu^2$ for any $r \varepsilon \Delta_{n\nu}$, $r \subset R_{n\nu}$, $\nu = 1, 2, \cdots, n$, $n = 1, 2, \cdots$. Let us now suppose $r \subset Q - R_{n\nu}$; then $a(\Sigma) \leq (1/m)\mathfrak{I}(\Sigma) = (1/m)[\mathfrak{I}(S_n) - \mathfrak{I}(S_{n\nu}')] \leq 3\mu_\nu/m$, and, by no. 27, $a(\Sigma) \leq (3/m)(2^{-2}M\epsilon_\nu^2) < (5M/m)\epsilon_\nu^2$ for any $r \varepsilon \Delta_{n\nu}$, $r \subset Q - R_{n\nu}$, $\nu = 1, 2, \cdots, n$, $n = 1, 2, \cdots$. The statement at the beginning of this no. is thus proved.

33. For each $n = 1, 2, \cdots$ and $\nu = 1, 2, \cdots, n$, let us apply Lemma 19 in connection with each rectangle $r \varepsilon \Delta_{n\nu}$, the vector $x_n(w)$, $w \varepsilon Q$, the constant $K = M/m \geq 1$, any point $x_0 = x_n(w_0)$, where $w_0 \varepsilon r^*$, and $D = 4\epsilon_\nu$.

This is possible because from no. 30 we have $\operatorname{diam} x_n(r^*) \leq 4\epsilon_\nu = D$ for any $r \varepsilon \Delta_{n\nu}$. Let $T_{n\nu}$ denote the collection of simple polygons $\pi \subset r$ defined by Lemma 19 in all rectangles $r \varepsilon \Delta_{n\nu}$. According to Lemma 19 all polygons π are interior to r together with their boundaries. Let $x_{n\nu}(w)$ be the vector defined by Lemma 19 on each rectangle $r \varepsilon \Delta_{n\nu}$. Because of $x_{n\nu}(w) = x_n(w)$ for any $w \varepsilon r^*$, as well as for any $w \varepsilon r - \Sigma\pi$, the vectors $x_{n\nu}(w)$, $w \varepsilon r$, $r \varepsilon \Delta_{n\nu}$ define a unique vector $x_{n\nu}(w)$, $w \varepsilon Q$, continuous and quasi-linear on Q. Let $S_{n\nu}$ be the polyhedral surface $S_{n\nu}: x = x_{n\nu}(w)$, $w \varepsilon Q$.

For each rectangle $r \varepsilon \Delta_{n\nu}$ let σ, σ_0 be the family of polyhedral surfaces defined by the vectors $x_n(w)$ and $x_{n\nu}(w)$ on the family of polygons π, $\pi \varepsilon T_{n\nu}$, $\pi \subset r$. We have $\partial\sigma = \partial\sigma_0$. Let $\mathfrak{L}$ be the total length of $\partial\sigma = \partial\sigma_0$, i. e., the sum of the lengths p of the common boundaries curves of the polyhedral surfaces defined by $x_n(w)$, $x_{n\nu}(w)$ on the polygons π, $\pi \varepsilon T_{n\nu}$. $\pi \subset r$, $r \varepsilon \Delta_{n\nu}$. We have $\mathfrak{L} = \Sigma p$, where the sum ranges over the polygons $\pi \varepsilon T_{n\nu}$, $\pi \subset r$.

34. Since $x_{n\nu}(w) = x_n(w)$ for any $w \varepsilon Q^*$ we have $\partial S_{n\nu} = \partial S_n$, $\| \partial S_{n\nu}, C \| = \| \partial S_n, C \| = \| C_n, C \| \leq 2\delta_n$. Each surface σ_0 is by Lemma 19 contained in the minimum convex body containing $[\mathfrak{L}] = [\partial\sigma] = [\partial\sigma_0]$ and, because of $[\mathfrak{L}] \subset [S_n] \subset A_{2\delta_n}$ where $A_{2\delta_n}$ is a convex body, we have $[S_{n\nu}] \subset A_{2\delta_n}$, $\nu = 1, 2, \cdots, n$. This assures us (no. 27) that $S_{n\nu} \varepsilon W_{2\delta_n}$, hence $\mathfrak{J}(S_{n\nu}) \geq j(2\delta_n) \geq j - \mu_n$, $\nu = 1, 2, \cdots, n$.

By Lemma 19 we have

$$a(\sigma_0) \leq (K^{-1}/4)a(\sigma) = (m/4M)a(\sigma).$$

By (1) (no. 26) it follows that

$$0 \leq \mathfrak{J}(\sigma_0) \leq Ma(\sigma_0) \leq M(m/4M)a(\sigma) \leq M(m/4M)(1/m)\mathfrak{J}(\sigma),$$

hence

$$0 \leq \mathfrak{J}(\sigma_0) \leq (1/4)\mathfrak{J}(\sigma). \tag{2}$$

Let us indicate by $S_{n\nu 0}$ the polyhedral surface (sum of triangles) defined by the vectors $x_{n\nu}(w)$, $x_n(w)$ on the polygonal region $Q - \Sigma\pi$, $\pi \varepsilon T_{n\nu}$. We have $S_{n\nu 0} + \Sigma\sigma = S_n$, $S_{n\nu 0} + \Sigma\sigma_0 = S_{n\nu}$, where the sum ranges on all families of surfaces σ [resp. σ_0], each family determined by the vector x_n [resp. $x_{n\nu}$] on the polygons π, $\pi \subset T_{n\nu}$, $\pi \subset r$, for any $r \varepsilon \Delta_{n\nu}$. We have

$$\mathfrak{J}(S_{n\nu 0}) + \sum_{r \varepsilon \Delta_{n\nu}} \mathfrak{J}(\sigma) = \mathfrak{J}(S_n), \qquad \mathfrak{J}(S_{n\nu 0}) + \sum_{r \varepsilon \Delta_{n\nu}} \mathfrak{J}(\sigma_0) = \mathfrak{J}(S_{n\nu}); \tag{3}$$

and, by (2), we have $\mathfrak{J}(S_{n\nu}) \leq \mathfrak{J}(S_n)$; hence $\mathfrak{J}(S_{n\nu}) \leq \mathfrak{J}(S_n) \leq j + 2\mu_n$, $\nu = 1, 2, \cdots, n$. We have thus proved that

$$j - \mu_n \leq \mathfrak{J}(S_{n\nu}) \leq j + 2\mu_n, \qquad \nu = 1, 2, \cdots, n.$$

By (2) and (3) we have also

$$3\mu_n \geq \mathfrak{F}(S_n) - \mathfrak{F}(S_{n\nu}) = \sum_{r \varepsilon \Delta_{n\nu}} [\mathfrak{F}(\sigma) - \mathfrak{F}(\sigma_0)]$$
$$\geq \sum_{r \varepsilon \Delta_{n\nu}} (1 - 1/4)\mathfrak{F}(\sigma) - (3/4) \sum_{r \varepsilon \Delta_{n\nu}} \mathfrak{F}(\sigma),$$

hence

$$\sum_{r \varepsilon \Delta_{n\nu}} \mathfrak{F}(\sigma) \leq 4\mu_n, \qquad \nu = 1, 2, \cdots, n. \tag{4}$$

If we denote by $\sigma_{n\nu}$, $\sigma_{n\nu 0}$ the total families of all surfaces σ, σ_0 for all $r \varepsilon \Delta_{n\nu}$, we have also, by (2) and (4),

$$\mathfrak{F}(\sigma_{n\nu 0}) \leq (1/4)\mathfrak{F}(\sigma_{n\nu}) \leq \mu_n, \qquad \nu = 1, 2, \cdots, n.$$

35. By (1) and Lemma 19 we have

$$a(\sigma) \leq (1/m)\mathfrak{F}(\sigma), \qquad \mathfrak{L}^2 \leq (1/K)a(\sigma) = (m/M)a(\sigma), \tag{5}$$

where $\mathfrak{L} = \partial\sigma = \partial\sigma_0$ and we have denoted by $\mathfrak{L}$ also the length of $\mathfrak{L}$. From (4) and (5) we have

$$\sum_{r \varepsilon \Delta_{n\nu}} \mathfrak{L}^2 \leq (m/M) \sum_{r \varepsilon \Delta_{n\nu}} a(\sigma) \leq (1/M) \sum_{r \varepsilon \Delta_{n\nu}} \mathfrak{F}(\sigma) \leq (1/M) 4\mu_n,$$

hence

$$\sum_{r \varepsilon \Delta_{n\nu}} \mathfrak{L}^2 \leq 4\mu_n/M, \qquad \sum_{r \varepsilon \Delta_{n\nu}} a(\sigma) \leq 4\mu_n/m;$$

and also, by no. 27,

$$\sum_{r \varepsilon \Delta_{n\nu}} \Big(\sum_{\substack{\pi \varepsilon T_{n\nu} \\ \pi \subset r}} p \Big)^2 = \sum_{r \varepsilon \Delta_{n\nu}} \mathfrak{L}^2 \leq 4\mu_n/M \leq \epsilon_n^2 < \epsilon_n, \qquad \nu = 1, 2, \cdots, n.$$

In particular, for each rectangle $r \varepsilon \Delta_{n\nu}$ the total length $\mathfrak{L}$ of the common boundary $\mathfrak{L} = \partial\sigma = \partial\sigma_0$ of the surfaces defined by the vectors $x_n(w)$, $x_{n\nu}(w)$ on the polygons π, $\pi \varepsilon T_{n\nu}$, $\pi \subset r$, satisfies the relation

$$\mathfrak{L} \leq [4\mu_n/M]^{1/2} \leq \epsilon_n, \qquad \nu = 1, 2, \cdots, n.$$

For any $\nu = 1, 2, \cdots, n$, we have $\nu \leq n$, $\epsilon_\nu \geq \epsilon_n$, hence $\mathfrak{L} \leq \epsilon_\nu$. For any polygon $\pi \varepsilon T_{n\nu}$ we have $\pi \subset r$, $r \varepsilon \Delta_{n\nu}$ and the diameter of the set $x_n(\pi^*)$ is $\leq$ the length of the curve $x = x_n(w)$, $w \varepsilon \pi^*$. But this curve is the boundary curve of one of the surfaces which are part of σ. Hence diam $x_n(\pi^*) \leq$ length $(\partial\sigma) = \mathfrak{L} \leq \epsilon_n \leq \epsilon_\nu$, $\nu = 1, 2, \cdots, n$.

36. For any $r \varepsilon \Delta_{n\nu}$ let us indicate by Σ, Σ_0 the polyhedral surfaces

$\Sigma: x = x_n(w),\ w \in r,\ \Sigma_0: x = x_{n\nu}(w),\ w \in r$. Then Σ_0 (Lemma 19) is completely contained in a sphere of center $x_0 = x_n(w_0),\ w_0 \in r^*$, and radius $\rho \leq 2D + 3[Ka(\Sigma)]^{1/2}$ whence, by no. 32, $\rho \leq 8\epsilon_\nu + 3[(M/m)(5M/m)\epsilon_\nu^2]^{1/2} < 15K\epsilon_\nu$. Therefore $\operatorname{diam} x_{n\nu}(r) = \operatorname{diam}(\Sigma_0) < 30K\epsilon_\nu$ for any $r \in \Delta_{n\nu}$, $\nu = 1, 2, \cdots, n,\ n = 1, 2, \cdots,\ K = M/m \geq 1$.

37. Let $T'_{n1} \subset T_{n1}$ be the subcollection of all polygons $\pi \in T_{n1}$ which are not completely contained in any rectangle $r \in \Delta_{n2} + \Delta_{n3} + \cdots + \Delta_{nn}$; let $T'_{n\nu} \subset T_{n\nu}$, $\nu = 2, 3, \cdots, n-1$, be the subcollection of all polygons $\pi \in T_{n\nu}$ which are not completely contained in any rectangle $r \in \Delta_{n,\nu+1} + \Delta_{n,\nu+2} + \cdots + \Delta_{nn}$ and not completely contained in the set $T'_{n1} + T'_{n2} + \cdots + T'_{n,\nu-1}$; let $T'_{nn} \subset T_{nn}$ be the subcollection of all polygons $\pi \in T_{nn}$ which are not completely contained in $T'_{n1} + T'_{n2} + \cdots + T'_{n,n-1}$.

Let us denote by $T_{n\nu}$, $T'_{n\nu}$ also the (closed) sets of points of Q covered by the polygons $\pi \in T_{n\nu}$, $T'_{n\nu}$. Let us denote by $\Lambda_{n\nu} \subset \Delta_{n\nu}$ the subcollection of all rectangles $r \in \Delta_{n\nu}$ which are not completely contained in the closed set $T'_{n1} + T'_{n2} + \cdots + T'_{n,\nu-1}$, $\nu = 1, 2, \cdots, n$. For any polygon $\pi \in T'_{n\nu}$ we have $\pi \subset r$, $r \in \Delta_{n\nu}$. We prove the following statement: π *is completely exterior to the set* $r(T'_{n1} + T'_{n2} + \cdots + T'_{n,\nu-1})$. First let us observe that $\pi \in T'_{n\nu}$, hence π is not completely contained in $T'_{n1} + T'_{n2} + \cdots + T'_{n,\nu-1}$. Here $\pi \subset r$, hence also r is not completely contained in $T'_{n1} + T'_{n2} + \cdots + T'_{n,\nu-1}$. Let π' be any polygon $T'_{n1} + \cdots + T'_{n,\nu-1}$; then r is not completely contained in π'. On the other hand, by definition of $T'_{n1}, T'_{n2}, \cdots, T'_{n,\nu-1}$, π' is not completely contained in r, because $r \in \Delta_{n\nu}$. If r and π' have no interior points in common the statement is trivial; if r and π' have interior points in common then they must have also points of their boundaries in common, that is $\pi'^* r^* \neq 0$ and each component of $\pi'^* r$ is a polygonal line having both end points on r^*. From no. 35 we have $\operatorname{diam} x_n(\pi'^*) \leq \epsilon_n \leq \epsilon_\nu < D$. By Lemma 19, statement v) we deduce that π'^* is completely outside of π. Because π is in r but not completely contained in π', it follows that π is outside of π'. This holds for all polygons $\pi' \in T'_{n1} + T'_{n2} + \cdots + T'_{n,\nu-1}$; hence π is outside of $r(T'_{n1} + T'_{n2} + \cdots + T'_{n,\nu-1})$ and the statement is proved.

38. By no. 37 no polygon $\pi \in T_{n\nu}'$ has points in common with polygons $\pi' \in T_{n1}' + \cdots + T_{n,\nu-1}'$, $\nu = 2, 3, \cdots, n$. Let us denote by $X_n(w)$, $w \in Q$, the vector which is $= x_{n\nu}(w)$ for any $w \in \pi$, $\pi \in T_{n\nu}'$, $\nu = 1, 2, \cdots, n$, and which is $= x_n(w)$ for any $w \in Q - I_n'$, $I_n' = T_{n1}' + \cdots + T_{nn}'$. The vector $X_n(w)$, $w \in Q$, is continuous and quasi-linear on Q. Let S_n be the polyhedral

surface $\bar{S}_n : x = X_n(w)$, $w \varepsilon Q$. The same reasoning as at the beginning of no. 34 assures us that $\partial\bar{S}_n = \partial S_n = C_n$, $\| \partial\bar{S}_n, C \| \leq 2\delta_n$, $[\bar{S}_n] \subset A_{2\delta_n}$, hence $\bar{S}_n \varepsilon W(2\delta_n)$ and $\Im(\bar{S}_n) \geq j - \mu_n$. The vector $X_n(w)$ defines on the polygons $\pi \varepsilon T_{n\nu}'$ a family $\sigma_{n\nu 0}'$ of polyhedral surfaces which, because of $T_{n\nu}' \subset T_{n\nu}$, is only a part of the family $\sigma_{n\nu 0}$ (no. 34), defined by $x_{n\nu}(w)$ on *all* polygons $\pi \varepsilon T_{n\nu}$ and which is the sum of the surfaces σ_0 of nos. 33-34 for all $r \varepsilon \Delta_{n\nu}$. If we denote by $\bar{S}_{n0}$ the surface (sum of triangles) defined by $X_n(w)$, as well as by $x_n(w)$, on $Q - I_n'$, then, by no. 28 and no. 34, we have

$$\Im(\bar{S}_n) = \Im(\bar{S}_{n0}) + \sum_{\nu=1}^{n} \Im(\sigma_{n\nu 0}') \leq \Im(S_n) + \sum_{\nu=1}^{n} \Im(\sigma_{n\nu 0})$$

$$\leq j + 2\mu_n + \sum_{\nu=1}^{n} \mu_n = j + (n+2)\mu_n.$$

By no. 27 we have finally

$$j - 2^{-n} < j - \mu_n \leq \Im(\bar{S}_n) \leq j + (n+2)\mu_n < j + 2^{-n}.$$

39. Let $\epsilon > 0$ be any real number. Let $\mu = \mu(\epsilon)$ be the first integer such that $132K\epsilon_\mu < \epsilon$. Let $\delta = \delta(\epsilon) = \eta_\mu > 0$. Let w_1, w_2 be any two points of Q with $| w_1 - w_2 | < \delta$ and let $n \geq \mu$ be any integer. Then $w_1 \varepsilon r_1$, $w_2 \varepsilon r_2$, $r_1, r_2 \varepsilon \Delta_{n\mu}$, where r_1, r_2 are the same rectangle of $\Delta_{n\mu}$, or two rectangles of $\Delta_{n\mu}$ having one side or one vertex in common. In any case $r_1^* r_2^* \neq 0$. We are going to consider several cases:

a) w_1, w_2 belong to the same polygon $\pi \varepsilon T_{n\nu}'$, $1 \leq \nu \leq n$. Then, by no. 38, $| X_n(w_1) - X_n(w_2) | = | x_{n\nu}(w_1) - x_{n\nu}(w_2) |$ and the last expression is $\leq$ the diameter of the minimum convex body containing $x_n(\pi^*)$, which is certainly $\leq 2 \operatorname{diam} x_n(\pi^*)$. Finally, by no. 35, we have

$$| X_n(w_1) - X_n(w_2) | < 2 \operatorname{diam} x_n(\pi^*) \leq 2\epsilon_n \leq 2\epsilon_\mu < \epsilon.$$

b) $w_i \varepsilon \pi_i$, $\pi_i \varepsilon T'_{n\nu_i}$, π_i not completely contained in r_i, $r_i \varepsilon \Delta_{n\mu}$, $i = 1, 2$, $\pi_1 \neq \pi_2$. Here $(r_1^* + r_2^*)\pi_i^*$ is not empty; hence let w_i' be any point $w_i' \varepsilon (r_1^* + r_2^*)\pi_i^*$, $i = 1, 2$. We have $| X_n(w_i) - X_n(w_i') | \leq 2\epsilon_n$, $i = 1, 2$. The two points w_1', w_2' are in $r_1^* + r_2^*$; hence there is a polygonal line λ joining w_1', w_2' contained in $r_1^* + r_2^*$, and therefore formed by at most four consecutive segments contained each in one side of the rectangles r_1, r_2. By no. 30 we have $\operatorname{diam} x_n(\lambda) \leq 4\epsilon_\mu$, hence $| x_n(w_1') - x_n(w_2') | \leq \operatorname{diam} x_n(\lambda)$ $\leq 4\epsilon_\mu$. On the other hand $X_n(w_i') = x_{n\nu_i}(w_i') = x_n(w_i')$, $i = 1, 2$. Therefore

$$| X_n(w_1) - X_n(w_2) | \leq | X_n(w_1) - x_n(w_1') | + | x_n(w_1') - x_n(w_2') |$$
$$+ | x_n(w_2') - X_n(w_2) | \leq 2\epsilon_n + 4\epsilon_\mu + 2\epsilon_n \leq 8\epsilon_\mu < \epsilon.$$

c) $w_i \,\varepsilon\, \pi_i$, $\pi_i \,\varepsilon\, T_{n\nu_i}'$, $\pi_i \subset r_i$, $r_i \,\varepsilon\, \Delta_{n\mu}$, $i = 1, 2$. Here π_i is completely contained in r_i, hence $n \geq \nu_i \geq \mu$ (no. 37). Let w_{i,ν_i+1} be any point $w_{i,\nu_i+1} \,\varepsilon\, \pi_i^*$, hence $| X_n(w_i) - X_n(w_{i,\nu_i+1}) | < 2\epsilon_n$, $X_n(w_{i,\nu_i+1}) = x_n(w_{i,\nu_i+1})$, $i = 1, 2$. We have $\pi_i \,\varepsilon\, T_{n\nu_i}'$, hence $\pi_i \subset r_{i\nu_i}$, $r_{i\nu_i} \,\varepsilon\, \Delta_{n\nu_i}$. For any point $w_{i\nu_i} \,\varepsilon\, r_{i\nu_i}^*$ we have $| x_n(w_{i,\nu_i+1}) - x_n(w_{i\nu_i}) | \leq \operatorname{diam} \Sigma_{\nu_i}$, where Σ_{ν_i} is the surface defined by application of Lemma 19 on the rectangle $r_{i\nu_i}$, hence (no. 36) $| x_n(w_{i,\nu_i+1}) - x_n(w_{i,\nu_i}) | \leq \operatorname{diam} \Sigma_{\nu_i} < 30K\epsilon_{\nu_i}$. Now $\pi_i \,\varepsilon\, T_{n\nu_i}'$, hence no polygon $\pi \,\varepsilon\, T_{n1}' + \cdots + T_{n,\nu_i-1}'$ (no. 37) has points in common with π_i. Therefore $r_{i\nu_i}$ is not completely contained in any of the polygons $\pi \,\varepsilon\, T_{n1}' + \cdots + T_{n,\nu_i-1}'$. In particular, $r_{i\nu_i}$ is not completely contained in any of the polygons $\pi \,\varepsilon\, T_{n,\nu_i-1}'$. The polygons $\pi \,\varepsilon\, T_{n,\nu_i-1} - T_{n,\nu_i-1}'$ are either completely contained in rectangles $r \,\varepsilon\, \Delta_{nt}$, $t \geq \nu_i$, or completely contained in $T_{n1}' + \cdots + T_{n,\nu_i-1}'$. Therefore $r_{i\nu_i}^*$ cannot be completely contained in the interior of any such polygons π. In any case we can choose on $r_{i\nu_i}^*$ the point $w_{i\nu_i}$ not interior to any of the polygons $\pi \,\varepsilon\, T_{n,\nu_i-1}$. The point $w_{i\nu_i}$ now chosen belongs to a rectangle $r_{i,\nu_i-1} \,\varepsilon\, \Delta_{n,\nu_i-1}$, and for any point $w_{i,\nu_i-1} \,\varepsilon\, r_{i,\nu_i-1}^*$ we have $| x_n(w_{i,\nu_i}) - x_n(w_{i,\nu_i-1})| \leq \operatorname{diam} \Sigma_{\nu_i-1} < 30K\epsilon_{\nu_j-1}$, where Σ_{ν_i-1} is defined as Σ_{ν_i}. By repeating this procedure we have a finite chain of points $w_{it} \,\varepsilon\, r_{it}^*$, $t = \nu_i$, $\nu_i - 1, \cdots, \mu + 1, \mu$, and $| x_n(w_{i,t+1}) - x_n(w_{it}) | < 30K\epsilon_t$, $t = \nu_i, \nu_i - 1, \cdots, \mu$, $i = 1, 2$. Therefore $| x_n(w_{i,\nu_i+1}) - x_n(w_{i\mu}) | \leq \Sigma \, | x_n(w_{it}) - x_n(w_{i,t+1}) | < 30K(\epsilon_\mu + \epsilon_{\mu+1} + \cdots + \epsilon_{\nu_i})$ and, because $\epsilon_{t+1} < \epsilon_t/2$, also $< 60K\epsilon_\mu$. Finally, if we denote by w_i', w_i'' the points w_{i,ν_i+1}, $w_{i\mu}$, we have

$$| X_n(w_i) - x_n(w_i') | < 2\epsilon_n \leqq 2\epsilon_\mu, \quad | x_n(w_i') - x_n(w_i'') | < 60K\epsilon_\mu, \quad w_i'' \,\varepsilon\, r_{i\mu},$$
$$i = 1, 2.$$

Let us now observe that the points $w_{i,t+1}$, w_{it} belong to the same rectangle $r_{it} \,\varepsilon\, \Delta_{nt}$, $(t = \nu_i, \nu_i - 1, \cdots, \mu)$, hence the distance $| w_{i,t+1} - w_{it} |$ is $<$ twice the maximum dimension of all rectangles $r \,\varepsilon\, \Delta_{nt}$, i. e., by no. 30, $| w_{i,t+1} - w_{it} | < 4\xi_t$, where $4\xi_t \leq (1/2)\eta_{t-1}$, $(1/2^2)\eta_{t-2}, \cdots$. Therefore

$$| w_i' - w_{i,\mu+1} | = | w_{i,\nu_i+1} - w_{i,\mu+1} | \leq \Sigma \, | w_{i,t+1} - w_{it} |,$$
$$t = \nu_i, \nu_i - 1, \cdots, \mu + 1,$$

and finally

$$| w_i' - w_{i,\mu+1} | < 4\xi_{\nu_i} + 4\xi_{\nu_i-1} + \cdots + 4\xi_{\mu+1}$$
$$\leq (1/2)\eta_\mu + (1/2^2)\eta_\mu + \cdots + (1/2^{\nu_i-\mu})\eta_\mu < \eta_\mu.$$

This assures that the point $w_{i,\mu+1}$ has a distance from $w_i' \varepsilon r_i[r_{i\mu}, r_i \varepsilon \Delta_{n\mu}]$ which is less than the minimum dimension of all rectangles $r \varepsilon \Delta_{n\mu}$. Also the point $w_i'' = w_{i\mu}$ belongs to the rectangle $r_i' \equiv r_{i\mu}$ and necessarily the rectangle r_i' is either coincident with r_i, or adjacent to r_i, or has a vertex in common with r_i, $i = 1, 2$. Therefore the rectangles r_1', r_2' belong to a chain r_1', r_1, r_2, r_2' of rectangles $r \varepsilon \Delta_{n\mu}$, any two consecutive ones being identical, or adjacent, or with a vertex in common. Consequently the points $w_1'' \varepsilon r_1'^*$, $w_2'' \varepsilon r_2'^*$ can be joined by a polygonal line λ formed by at most eight consecutive segments each contained in a side of a rectangle $r \varepsilon \Delta_{n\mu}$. This implies that

$$| x_n(w_1'') - x_n(w_2'') | \leq \operatorname{diam} x_n(\lambda) \leq 8\epsilon_\mu.$$

Finally we have

$$\begin{aligned} | X_n(w_1) - X_n(w_2) | &\leq | X_n(w_1) - x_n(w_1') | + | x_n(w_1') - x_n(w_1'') | \\ &\quad + | x_n(w_1'') - x_n(w_2'') | + | x_n(w_2'') - x_n(w_2') | + | x_n(w_2') - X_n(w_2) | \\ &\leq 2\epsilon_\mu + 60K\epsilon_\mu + 8\epsilon_\mu + 60K\epsilon_\mu + 2\epsilon_\mu \leq 132K\epsilon_\mu < \epsilon. \end{aligned}$$

d) $w_i \varepsilon Q - I_n'$, $i = 1, 2$. Here w_i not $\epsilon I_n'$, $I_n' = T_{n1}' + T_{n2}' + \cdots + T_{nn}'$, hence w_i not $\epsilon (T_{n1}' + \cdots + T_{n,n-1}')$. If w_i belonged to a polygon π, $\pi \varepsilon T_{nn} - T_{nn}'$ (as an interior or boundary point), then necessarily the polygon π would not be completely contained in the closed set $T_{n1}' + T_{n2}' + \cdots + T_{n,n-1}'$, hence $\pi \varepsilon T_{nn}'$ (no. 37), $\pi \subset I_n'$, which is impossible. This proves that w_i not ϵT_{nn}. Let us indicate by μ_i, $n \geq \mu_i \geq \mu$, the minimum of the indices t such that w_i not ϵT_{nt}, T_{nt} closed, $t = n, n-1, \cdots, \mu_i$. Then w_i not $\epsilon T_{n\mu_i}$, $w_i \varepsilon r_{i\mu_i}$, $i = 1, 2$. We have $X_n(w_i) = x_n(w_i)$, $i = 1, 2$. We can repeat the reasoning of c) with the simplification due to the fact that here $w_i' = w_i$, $i = 1, 2$. We obtain

$$\begin{aligned} | X_n(w_1) - X_n(w_2) | = | x_n(w_1) - x_n(w_2) | &\leq | x_n(w_1) - x_n(w_1'') | \\ &\quad + | x_n(w_1'') - x_n(w_2'') | + | x_n(w_2'') - x_n(w_2) | \\ &< 60K\epsilon_\mu + 4\epsilon_\mu + 60K\epsilon_\mu \leq 124K\epsilon_\mu < \epsilon. \end{aligned}$$

e) The two points w_i are in two different ones of the situations discussed in b), c), d). We have only to combine the different reasonings.

40. We have proved in the previous no. 39 that, given $\epsilon > 0$, there is an integer $\mu = \mu(\epsilon) > 0$ and a real number $\delta = \delta(\epsilon) > 0$ such that, for any two points $w_1, w_2 \varepsilon Q$, $| w_1 - w_2 | < \delta$, and for any integer $n \geq \mu$, we have

$|X_n(w_1) - X_n(w_2)| < \epsilon$. Now the $\mu - 1$ vectors $X_n(w)$, $w \varepsilon Q$, $n = 1, 2, \cdots, \mu - 1$, are continuous on Q, hence there is a number $\delta' = \delta'(\epsilon) > 0$ such that for any $w_1, w_2 \varepsilon Q$, $|w_1 - w_2| < \delta'$, and any $n = 1, 2, \cdots, \mu - 1$, we have $|X_n(w_1) - X_n(w_2)| < \epsilon$. If $\delta_0 = \delta_0(\epsilon) = \min[\delta, \delta']$, we have $|X_n(w_1) - X_n(w_2)| < \epsilon$ for any $w_1, w_2 \varepsilon Q$, $|w_1 - w_2| < \delta_0$ and any $n = 1, 2, \cdots,$. We have proved that the vectors $X_n(w)$, $w \varepsilon Q$, $n = 1, 2, \cdots,$ are *equicontinuous* on Q. From $[\bar{S}_n] \subset A_{2\delta_n}$, $n = 1, 2, \cdots$ (no. 36), it follows that the vectors $X_n(w)$, $w \varepsilon Q$, $n = 1, 2, \cdots$, are also *equibounded* in Q.

41. By Ascoli's theorem there exists a subsequence $X_{n_m}(w)$ of the vectors $X_n(w)$, $n = 1, 2, \cdots$, which converges uniformly toward a continuous vector $X(w)$, $w \varepsilon Q$. For simplicity's sake we suppose $n_m = m$. Such a vector coincides, on Q^*, with the vector $X(w)$ already determined in no. 29. Let S be the surface $S: x = X(w)$, $w \varepsilon Q$. From $X_n(w) \rightrightarrows X(w)$, $w \varepsilon Q$, and $\bar{S}_n: x = X_n(w)$, $w \varepsilon Q$, it follows that $\| \bar{S}_n, S \| \to 0$ and, by no. 29, also $\partial S = C$. We have $\| \bar{S}_n, S \| \to 0$ and, by no. 38, $a(\bar{S}_n) < (1/m)\mathfrak{J}(\bar{S}_n) < (1/m)(j+1) < +\infty$, hence $0 \leq L(S) \leq \lim a(\bar{S}_n) < +\infty$. From $[\bar{S}_n] \subset A_{2\delta_n}$, $\delta_n \to 0$, it follows that $[S] \subset A$, where A is a closed convex set of E_3. By no. 27 we have $S \subset W$, hence $\mathfrak{J}(S) \geq i$. Now we have $\delta_n \leq 1$, $n = 1, 2, \cdots$, hence all surfaces $S, \bar{S}_n$ are contained in the finite closed set A_2. We have $F_0(x, p) > 0$ for any $x \varepsilon A_2$, $|p| \neq 0$. The surface S is contained in A, hence in A_2, and the function $F_0(x, p)$ has continuous first derivatives with respect to p^s, $s = 1, 2, 3$, for all $x \varepsilon A$, $|p| \neq 0$ [because $F = F_0$ for $x \varepsilon A$], hence $F_0(x, p)$ has continuous first derivatives also for all $x \varepsilon [S]$, $|p| \neq 0$. We know that $\mathfrak{E}(x, p, \bar{p}) \geq 0$ for all $p \neq \bar{p}$, $|p|$, $|\bar{p}| \neq 0$, $x \varepsilon A$, hence also for all $x \varepsilon [S]$. All conditions of Lemma 9 (note in no. 8), where A is now the set A_2, are satisfied. Therefore the integral

$$\mathfrak{J}(S) = \int\int_S F_0(x, p)\, dw$$

is lower semicontinuous on the surface S with respect to the surfaces $\bar{S}_n$, $n = 1, 2, \cdots$, i.e. $\mathfrak{J}(S) \leq \underline{\lim}\, \mathfrak{J}(\bar{S}_n)$, where, by no. 38, we know that $\mathfrak{J}(\bar{S}_n) < j + 2^{-n}$, $n = 1, 2, \cdots$. Hence $\mathfrak{J}(S) \leq j$ and finally $i \leq \mathfrak{J}(S) \leq j$, where, by no. 27, we know that $j \leq i$. This implies that $\mathfrak{J}(S) = i = j$. We have thus proved that there exists a surface $S \varepsilon W$ for which $\mathfrak{J}(S)$ assumes its absolute minimum i in W. Theorem I (no. 25) is completely proved.

42. THEOREM II. *Let C be a closed Jordan curve in E_3, W the family of all oriented Fréchet surfaces S in E_3, whose Lebesgue area is finite and whose boundary curve is C, and let us suppose W is not empty. Let $\mathfrak{J}(S)$ be any positive definite semiregular integral such that $m\,|p| \leq F(x, p) \leq M\,|p|$, where $m, M, 0 < m \leq M < +\infty$, are two given constants, $x \,\varepsilon\, E_3$, $|p| \neq 0$. Then $\mathfrak{J}(S)$ has an absolute minimum in W.*

The proof is the same as for Theorem I with the obvious simplification in no. 26. Here we have $A = E_3$. From the equicontinuity of the vectors $X_n(w)$ in Q and the fact that we have $\| C_n, C \| \to 0$, it follows that the same vectors $X_n(w)$ are also equibounded in Q and therefore we can apply, here also, Ascoli's theorem.

43. *Note.* In the proofs of Theorems I and II the principle of Dirichlet is not applied directly. We observe however that Lemmas 5, 6, 12 have been applied, that the proof of Lemma 6 is based on Lemma 5 and that the recalled proofs of Lemmas 5 and 12 follow reasonings which are the same as for some existence theorems for the problem of Dirichlet. Nevertheless our proofs appear to need somewhat weaker facts than the principle of Dirichlet, namely only that a) any non-degenerate surface of finite Lebesgue area posseses Dirichlet representations; b) the g. l. b. of the relative Dirichlet integrals coincides with the Lebesgue area, without stating that the minimum is actually attained.

THE UNIVERSITY OF BOLOGNA AND
PURDUE UNIVERSITY.

BIBLIOGRAPHY.

[1] L. Cesari, "Criteri di uguale continuità ed applicazioni alla quadratura delle superficie," *Annali Scuola Normale Superiore Pisa*, vol. 12 (1943), pp. 61-84.

[2] ———, "La nozione di integrale sopra una superficie in forma parametrica," *Ibid.*, vol. 13 (1947), pp. 73-117.

[3] ———, "Condizioni sufficienti per la semicontinuità degli integrali sopra una superficie in forma parametrica," *Ibid.*, vol. 14 (1948), pp. 47-79.

[4] ———, "Condizioni necessarie per la semicontinuità degli integrali sopra una superficie in forma parametrica," *Annali di Matematica*, vol. 29 (1950), pp. 199-224.

[5] ———, "Area and representation of surfaces," *Bulletin of the American Mathematical Society*, vol. 56 (1950), pp. 218-232.

[6] ———, "Sulle superficie di Fréchet," *Rivista di Matematica dell'Università di Parma*, vol. 1 (1950), pp. 19-44.

[7] ———, "Su un particolare processo di retrazione per superficie," *Ibid.*, vol. 2 (1951).

[8] ———, "Problemi di Calcolo delle Variazioni e questioni connesse," *Ibid.*, vol. 1 (1950), pp. 293-304.

[9] S. Cinquini, "Sull'approssimazione delle funzioni di due variabili," *Annali di Matematica*, vol. 11 (1933), pp. 295-323.

[10] R. Courant, "On the problem of Plateau," *Proceedings of the National Academy of Science*, vol. 22 (1936), pp. 367-372.

[11] J. Douglas, "Solution of the problem of Plateau," *Transactions of the American Mathematical Society*, vol. 33 (1931), pp. 263-321.

[12] S. Eilenberg, "On Φ measures," *Annales Société Polonaise de Mathématique*, vol. 17 (1938), pp. 252-253.

[13] ———, (with O. E. Harrold) "Continua of finite linear measure," *American Journal of Mathematics*, vol. 65 (1943), pp. 137-146.

[14] E. J. McShane, "On the semicontinuity of integrals in the Calculus of Variations," *Annals of Mathematics*, vol. 33 (1932), pp. 460-484.

[15] ———, "Integrals over surfaces in parametric form," *Annals of Mathematics*, vol. 34 (1933), pp. 815-838.

[16] ———, "Parametrizations of saddle surfaces, with applications to the problem of Plateau," *Transactions of the American Mathematical Society*, vol. 35 (1933), pp. 716-733.

[17] ———, "Existence theorems for double integral problems of the Calculus of Variations," *Ibid.*, vol. 38 (1935), pp. 549-563.

[18] ———, "Extension of range of functions," *Bulletin of the American Mathematical Society*, vol. 40 (1934), pp. 837-842.

[19] C. B. Morrey, "A class of representation of manifolds," *American Journal of Mathematics*, vol. 55 (1933), pp. 683-707; vol. 56 (1934), pp. 275-293.

[20] ———, "The topology of (path) surfaces," *Ibid.*, vol. 57 (1935), pp. 17-50.

[21] ———, "An analytic characterization of surfaces of finite Lebesgue area," *Ibid.*, vol. 57 (1935), pp. 692-702; vol. 58 (1936), pp. 313-322.

[22] T. Rado, "On the problem of Plateau," *Ergebnisse der Mathematik und ihrer Grenzgebiete*, vol. 2 (1933).

[23] ———, "On the semicontinuity of double integrals in the parametric form," *Transactions of the American Mathematical Society*, vol. 51 (1942), pp. 336-361.

[24] ———, "Length and Area," *Colloquium Publications of the American Mathematical Society*, vol. 30 (1948).

[25] P. V. Reichelderfer, "On the semicontinuity of double integrals," *Transactions of the American Mathematical Society*, vol. 67 (1949), pp. 275-285.

[26] L. Tonelli, *Fondamenti di Calcolo delle Variazioni*, Bologna, Zanichelli (1923-1924).

[27] ———, "Sulla rappresentazione analitica delle funzioni di variabili reali," *Rendiconti Circolo Matematico di Palermo*, vol. 29 (1910), pp. 1-36.

[28] ———, "Sur la semicontinuité des integrales doubles du Calcul des Variations," *Acta Mathematica*, vol. 53 (1929), pp. 325-346.

[29] ———, "L'estremo assoluto degli integrali doppi," *Annali Scuola Normale Superiore Pisa*, vol. 2 (1933), pp. 89-130.

[30] ———, "Su l'esistenze delle estremali assolute per gli integrali doppi," *Ibid.*, vol. 8 (1939), pp. 161-165.

[31] ———, "Sul problema di Plateau," *Rendiconti Accademia dei Lincei*, vol. 29 (1936), pp. 333-339, 393-398.

[32] L. C. Young, "A lemma in the theory of surfaces," *Journal of the London Mathematical Society*, vol. 19 (1944), pp. 209-219.

[33] ———, "On the isoperimetric ratio for a harmonic surface," *Proceedings of the London Mathematical Society*, vol. 49 (1947), pp. 396-408.

[34] ———, "On area and length," *Fundamenta Mathematicae*, vol. 35 (1948), pp. 275-302.

[35] ———, "Some applications of the Dirichlet integral to the theory of surfaces," *Transactions of the American Mathematical Society*, vol. 64 (1948), pp. 317-335.

THE "INDEPENDENT SCALARS" IN HOMOGENEOUS TURBULENCE.*

By N. Coburn.

1. Introduction. In discussing my paper [1] at the University of Illinois Symposium on Fluid Dynamics, Professor S. Chandrasekhar mentioned the problem of determining a method for generating the "independent scalars" in various types of homogeneous turbulence. By definition, these scalars must possess three properties: (1) they must be components of the various correlation tensors; (2) these scalars must be defined in such a manner that the correlation tensors satisfy the continuity relations identically; (3) the maximum number of such scalars must occur in each correlation tensor.

It is the purpose of this paper to furnish a method for the construction of such independent scalars. The method is based on the following considerations. First, one determines the maximum number of independent scalars associated with a given correlation tensor. Secondly, by a contraction process, a new tensor, containing the desired number of independent scalars, is constructed. By forming the curl of this tensor, one obtains a tensor which has two properties: (1) its divergence is zero (and hence it satisfies the continuity relations); (2) it possesses the structure of a correlation tensor which depends upon the maximum number of independent scalars.

In order to develop the method, we consider the cases of: (1) isotropic turbulence; (2) turbulence possessing axial symmetry with respect to a specified direction. The expressions for the correlation tensors in terms of the independent scalars are well known for the first case from the work of Kármán-Howarth [2] and H. P. Robertson.[3] For the second case, S. Chandrasekhar [4] has derived, recently, the corresponding results for the correlation tensors.

* Received October 26, 1950.

[1] N. Coburn, "A method for constructing correlation tensors in homogeneous turbulence," *Proceedings of the University of Illinois Symposium on Fluid Dynamics*, to be published.

[2] T. von Kármán and L. Howarth, "Isotropic turbulence," *Proceedings of the Royal Society*, vol. 164 (1938), pp. 192-215.

[3] H. P. Robertson, "An invariant theory of isotropic turbulence," *Proceedings of the Cambridge Philosophical Society*, vol. 36 (1940), pp. 209-223.

[4] S. Chandrasekhar, "The theory of axisymmetric turbulence," *Philosophical Transactions of the Royal Society*, vol. 242 (1950), pp. 557-577.

Finally, we discuss the general theory for constructing correlation tensors in terms of independent scalars. It is shown that if we require the correlation tensors to be *linear functions* of the independent scalars and their derivatives then, essentially, the above method is the one and only method for constructing such tensors.

The author wishes to thank Professor Chandrasekhar for allowing the author to see the proofs of his papers on axisymmetric turbulence (see footnote 4) and for his remarks relating the results of this paper with those of his paper.

2. The independent scalars in isotropic turbulence. Here, we consider the theory of Kármán-Howarth and Robertson. From the work of either of these authors, it is known that the correlation tensors have simple forms. In order to introduce these formulas, we mention our notation. Let x^λ, $\lambda = 1, 2, 3$, denote a Cartesian orthogonal coördinate system in Euclidean three-space. Further, if $P(x^\lambda)$, $'P('x^\lambda)$ denote the two points involved in the construction of the correlation tensors, then $\xi^\lambda = {}'x^\lambda - x^\lambda$ will denote the vector joining these points. Evidently, the magnitude, r, of this vector is,

$$r = [(\xi^1)^2 + (\xi^2)^2 + (\xi^3)^2]^{\frac{1}{2}}.$$

One additional tensor, which enters into this theory, is the metric tensor, $g^{\lambda\mu}$. With the aid of this tensor and vector, ξ^λ, the correlation tensors may be written as

$$(2.1)\qquad \begin{aligned} R^\lambda &= A\xi^\lambda, \qquad R^{\lambda\mu} = B\xi^\lambda\xi^\mu + Cg^{\lambda\mu}, \\ R^{\lambda\mu\alpha} &= D\xi^\lambda\xi^\mu\xi^\alpha + E(\xi^\lambda g^{\mu\alpha} + \xi^\mu g^{\lambda\alpha}) + F\xi^\alpha g^{\lambda\mu}, \end{aligned}$$

where A, B, C, D, E, F are scalar functions of r.

These tensors satisfy the following continuity relations

$$(2.2)\qquad R^\lambda{}_{,\lambda} = R^{\lambda\mu}{}_{,\mu} = R^{\lambda\mu\alpha}{}_{,\alpha} = 0,$$

where commas denote differentiation, that is $R^{\lambda\mu}{}_{,\mu} = \partial R^{\lambda\mu}/\partial\xi^\mu$ etc. The relation (2.2) furnishes a first order differential equation for A. Hence R^λ does not depend upon an independent scalar. Further, we notice that (2.2) determines a vanishing correlation vector and a vanishing correlation tensor of the second order. That is, the above relations imply that B, C are related by a single first order differential equation and that D, E, F are related by two first order differential equations. Thus, both $R^{\lambda\mu}$ and $R^{\lambda\mu\alpha}$ depend upon one independent scalar. In the next two paragraphs, we shall determine expressions for $R^{\lambda\mu}$, $R^{\lambda\mu\alpha}$ which depend upon single independent scalars, respectively.

First, we consider the tensor $R^{\lambda\mu}$. Multiplying the second equation of (2.1) by $e_{\gamma\lambda\beta}\xi^{\beta}$, we obtain

$$\bar{S}_{\gamma}{}^{\cdot\mu} = e_{\gamma\lambda\beta}R^{\mu\lambda}\xi^{\beta} - Ce_{\gamma\lambda\beta}g^{\lambda\mu}\xi^{\beta}. \tag{2.3}$$

Note, $e_{\gamma\lambda\beta}$ is the permutation tensor density. If we restrict our coordinate systems to right handed Cartesian orthogonal coordinates then $e_{\gamma\lambda\beta}$ is a tensor. Next, we introduce the tensor

$$\bar{R}^{\mu\alpha} = e^{\alpha\gamma\rho}\bar{S}_{\gamma}{}^{\cdot\mu}{}_{,\rho}. \tag{2.4}$$

It is easily verified that the tensor $\bar{R}^{\mu\alpha}$ satisfies (2.2), identically. That is, this tensor has zero divergence. Further, from (2.3), we see that this tensor depends upon one scalar. It remains to be shown that $\bar{R}^{\mu\alpha}$ has the form (2.1) and hence can be considered as a correlation tensor. Substituting (2.3) into (2.4), we obtain

$$\bar{R}^{\mu\alpha} = -\delta_{\lambda\beta}{}^{\alpha\rho}g^{\lambda\mu}(C\delta_{\rho}{}^{\beta} + r^{-1}\xi^{\beta}\xi^{\gamma}g_{\rho\gamma}dC/dr), \tag{2.5}$$

where $\delta_{\beta}{}^{\alpha}$ is the Kronecker tensor and $\delta_{\lambda\beta}{}^{\alpha\rho}$ is the generalized Kronecker tensor. In particular, $\delta_{\lambda\beta}{}^{\alpha\rho}$ satisfies the equation

$$\delta_{\lambda\beta}{}^{\alpha\rho} = e^{\mu\alpha\rho}e_{\mu\lambda\beta} = \delta_{\lambda}{}^{\alpha}\delta_{\beta}{}^{\rho} - \delta_{\lambda}{}^{\rho}\delta_{\beta}{}^{\alpha}.$$

Through use of this relation, we may reduce (2.5) to the equation

$$\bar{R}^{\mu\alpha} = -(2C + rdC/dr)g^{\mu\alpha} + r^{-1}dC/dr\xi^{\alpha}\xi^{\mu}.$$

That is, the above expression furnishes *a second order correlation tensor which depends upon a single scalar and for which the continuity relation* (2.2) *is identically satisfied.* It should be noted that (2.5) does not assure us that $\bar{R}^{\mu\alpha}$ will be symmetric. Such symmetry can be obtained, if needed, by use of an additional procedure, to be explained in section 6.

Now, we turn to the study of the third order correlation tensor, $R^{\lambda\mu\alpha}$. Similarly to (2.3) we introduce the tensor

$$\bar{S}_{\gamma}{}^{\cdot\lambda\mu} = e_{\gamma\alpha\beta}R^{\lambda\mu\alpha}\xi^{\beta} = Ee_{\gamma\alpha\beta}(\xi^{\lambda}g^{\mu\alpha} + \xi^{\mu}g^{\lambda\alpha})\xi^{\beta}. \tag{2.6}$$

Further, we construct the tensor

$$\bar{R}^{\lambda\mu\nu} = e^{\nu\gamma\rho}\bar{S}_{\gamma}{}^{\cdot\lambda\mu}{}_{,\rho}. \tag{2.7}$$

Evidently, the divergence relation (2.2) is identically satisfied. Further, the tensor $\bar{R}^{\lambda\mu\alpha}$ depends upon a single scalar. Hence this scalar is an independent scalar. Substituting (2.6) into (2.7), we obtain

(2. 8)

$$R^{\lambda\mu\nu} = -\delta_{\alpha\beta}{}^{\nu\rho}[r^{-1}dE/dr g_{\delta\rho}\xi^{\delta}\xi^{\beta}(\xi^{\lambda}g^{\mu\alpha} + \xi^{\mu}g^{\lambda\alpha}) + E(\delta_{\rho}{}^{\lambda}g^{\mu\alpha}\xi^{\beta} + \delta_{\rho}{}^{\mu}g^{\lambda\alpha}\xi^{\beta} + \delta_{\rho}{}^{\beta}\xi^{\lambda}g^{\mu\alpha} + \delta_{\rho}{}^{\beta}\xi^{\mu}g^{\lambda\alpha})].$$

Simplifying (2. 8), we find that

(2. 9) $$R^{\lambda\mu\nu} = 2(r^{-1}dE/dr)\xi^{\lambda}\xi^{\mu}\xi^{\nu} - (3E + rdE/dr)(\xi^{\lambda}g^{\mu\nu} + \xi^{\mu}g^{\lambda\nu}) + 2E\xi^{\nu}g^{\lambda\mu}.$$

Thus, (2. 9) *furnishes a third order correlation tensor which depends upon a single scalar and for which the continuity relation* (2. 2) *is identically satisfied. That is, we have constructed a third order correlation tensor of the type given in* (2. 1), *and such that this tensor is expressed in terms of one independent scalar* (*the maximum number of such scalars*).

3. The independent scalar in turbulence with axial symmetry. The general theory of turbulence with axial symmetry has been initiatel by G. K. Batchelor [5] and completely studied by S. Chandrasekhar.[4] In addition, another approach to this theory has been given by the present author.[1] The previously mentioned authors have followed the invariant theory method of H. P. Robertson.[3] Here, we shall apply our present ideas to the work of S. Chandrasekhar.

Let j^{λ} $(\lambda = 1, 2, 3)$ denote a unit vector in the axial direction. By following the invariant theory approach, Chandrasekhar constructs the following correlation tensors:

(3. 1) $$R^{\lambda} = M\xi^{\lambda} + Nj^{\lambda},$$

(3. 2) $$R^{\lambda\mu} = A\xi^{\lambda}\xi^{\mu} + Bg^{\lambda\mu} + Cj^{\lambda}j^{\mu} + Dj^{\lambda}\xi^{\mu} + E\xi^{\lambda}j^{\mu},$$

(3. 3) $$R^{\lambda\mu\alpha} = F\xi^{\lambda}\xi^{\mu}\xi^{\alpha} + G(j^{\lambda}\xi^{\mu} + \xi^{\lambda}j^{\mu})\xi^{\alpha} + H(j^{\lambda}\xi^{\mu} + \xi^{\lambda}j^{\mu})j^{\alpha} + Ig^{\lambda\mu}\xi^{\alpha} + Jg^{\lambda\mu}j^{\alpha} + Kj^{\lambda}j^{\mu}j^{\alpha} + Lj^{\lambda}j^{\mu}\xi^{\alpha} + Q\xi^{\mu}\xi^{\lambda}j^{\alpha} + R(\xi^{\lambda}g^{\mu\alpha} + \xi^{\mu}g^{\lambda\alpha}) + S(j^{\lambda}g^{\mu\alpha} + j^{\mu}g^{\lambda\alpha}).$$

The scalars N, M, etc., are considered to be functions of r, and μ where

(3. 4) $$r = [(\xi^{1})^{2} + (\xi^{2})^{2} + (\xi^{3})^{2}]^{1/2}, \qquad r\mu = \xi^{\lambda}j_{\lambda}.$$

The problem is to find the correlation tensors in terms of the independent scalars for which the continuity relations (2. 2) will be identically satisfied. First, we note that (2. 2) furnishes a single scalar relation between the scalars

[5] G. K. Batchelor, "On the theory of axisymmetric turbulence," *Proceedings of the Royal Society*, vol. 186 (1946), pp. 480-490.

M, N of (3.1). Hence, only one of the scalars, M, N, is independent. Secondly, we note that the second continuity relation determines a correlation vector. Since the general correlation vector for axial symmetry is of the type (3.1), the equation (2.2) furnishes two relations for the coefficients A, B, C, D, E. That is, three of these scalars are independent. By a similar argument, we see that the ten coefficients F, G, etc., satisfy four relations. Hence, six independent scalars characterize the third order correlation tensor $R^{\lambda\mu\alpha}$.

In order to express the correlation vector, R^λ, in terms of a single independent scalar, we proceed as in section 2. Multiplying (3.1) by $e_{\gamma\lambda\beta}\xi^\beta$ we obtain

$$S_\gamma = e_{\gamma\lambda\beta}R^\lambda\xi^\beta = Ne_{\gamma\lambda\beta}j^\lambda\xi^\beta. \tag{3.5}$$

We shall show that the vector

$$\bar{R}^\alpha = e^{\alpha\gamma\rho}S_{\gamma,\rho} \tag{3.6}$$

is the desired vector. First, we note that the vector, $\bar{R}^\alpha$, has zero divergence. Secondly, we note that

$$\partial/\partial\xi^\rho = g_{\rho\nu}[\xi^\nu(r^{-1}\partial/\partial r - \mu r^{-2}\partial/\partial\mu) + j^\nu r^{-1}\partial/\partial\mu]. \tag{3.7}$$

Substituting (3.5) into (3.6), we obtain the formula

$$\bar{R}^\alpha = -\delta_{\lambda\beta}{}^{\alpha\rho}(j^\lambda\xi^\beta N_{,\rho} + Nj^\lambda\delta_\rho{}^\beta). \tag{3.8}$$

Simplifying (3.8) by use of the expansion formula for $\delta_{\lambda\mu}{}^{\alpha\rho}$ and (3.7) we find that

$$\bar{R}^\alpha = [\mu\partial N/\partial r + r^{-2}(1-\mu^2)\partial N/\partial\mu]\xi^\alpha - [r\partial N/\partial r + 2N]j^\alpha. \tag{3.9}$$

The above equation is essentially equation (25) in Chandrasekhar's paper.[4]

Next, we consider the tensor $R^{\lambda\mu}$. We noted in the previous paragraph that this tensor should be expressible in terms of three independent scalars. Hence, we repeat the technique used in section 2. That is, we introduce the tensor

$$S_\gamma{}^{\cdot\mu} = e_{\gamma\lambda\beta}R^{\lambda\mu}\xi^\beta = e_{\gamma\lambda\beta}(Bg^{\lambda\mu} + Cj^\lambda j^\mu + Dj^\lambda\xi^\mu)\xi^\beta. \tag{3.10}$$

With the aid of (3.10), we form the tensor $\bar{R}^{\mu\alpha}$ of (2.4) and obtain

(3.11)

$$\begin{aligned}\bar{R}^{\mu\alpha} = &-\delta_{\lambda\beta}{}^{\alpha\rho}\xi^\beta[g^{\lambda\mu}(r^{-1}\partial B/\partial r\xi_\rho - r^{-2}\mu\partial B/\partial\mu\xi_\rho + r^{-1}\partial B/\partial\mu j_\rho)\\ &+ j^\lambda j^\mu(r^{-1}\partial C/\partial r\xi_\rho - r^{-2}\mu\partial C/\partial\mu\xi_\rho + r^{-1}\partial C/\partial\mu j_\rho)\\ &+ j^\lambda\xi^\mu(r^{-1}\partial D/\partial r\xi_\rho - r^{-2}\mu\partial D/\partial\mu\xi_\rho + r^{-1}\partial D/\partial\mu j_\rho) + Dj^\lambda\delta_\rho{}^\mu]\\ &-\delta_{\lambda\beta}{}^{\alpha\rho}\delta_\rho{}^\beta[Bg^{\lambda\mu} + Cj^\lambda j^\mu + Dj^\lambda\xi^\mu].\end{aligned}$$

For the sake of simplicity, we have written ξ_ρ for $g_{\rho\gamma}\xi^\gamma$ in the above equation. Replacing $\delta_{\lambda\beta}{}^{\alpha\rho}$ by $(\delta_\lambda{}^\alpha\delta_\beta{}^\rho - \delta_\beta{}^\alpha\delta_\lambda{}^\rho)$ we find that (3.11) becomes

(3.12)

$$\begin{aligned}\bar{R}^{\mu\alpha} = {} & \xi^\alpha\xi^\mu[r^{-1}\partial B/\partial r - r^{-2}\mu\partial B/\partial\mu + \mu\partial D/dr + (1-\mu^2)\partial D/\partial\mu] \\ & - g^{\alpha\mu}(2B + r\partial B/\partial r) + \xi^\alpha j^\mu[r^{-1}\partial B/\partial\mu + \mu\partial C/\partial r \\ & + r^{-1}(1-\mu^2)\partial C/\partial\mu + D] - j^\alpha\xi^\mu(r\partial D/\partial r + 3D) \\ & - j^\alpha j^\mu(2C + r\partial C/\partial r).\end{aligned}$$

The tensor $\bar{R}^{\mu\alpha}$ of (3.12) does not coincide with the corresponding tensor developed by Chandrasekhar.[6] The reason for this difference lies in the non-uniqueness of A, B, etc. That is, another possible correlation tensor can be formed from linear combinations of A, B, etc. or from

$$(3.13) \qquad {}'\bar{S}_{\gamma\cdot}{}^{\mu} = e_{\gamma\lambda\beta}{}'R^{\lambda\mu}j^\beta = e_{\gamma\lambda\beta}(F\xi^\lambda\xi^\mu + Gg^{\lambda\mu} + H\xi^\lambda j^\mu)j^\beta,$$

where ${}'R^{\lambda\mu} = F\xi^\lambda\xi^\mu + Gg^{\lambda\mu} + H\xi^\lambda j^\mu + Lj^\lambda\xi^\mu + Mj^\lambda j^\mu$. By expressing F, G, H as linear combinations of B, C, D with variable coefficients and adding the resulting tensor ${}'\bar{R}^{\lambda\mu}$ to $\bar{R}^{\lambda\mu}$, one obtains other possible second order correlation tensors. The uniqueness problem will be discussed in the next section. To calculate the third order correlation tensors, one applies (2.6), (2.7) to the tensor (3.3). The calculations are routine but rather lengthy and will be omitted.

4. The theory of correlation tensors which depend upon independent scalars. In this section, we discuss the mathematical significance of our method.

a. *The General Problem.* We consider the following problem: *given the arbitrary correlation tensors $R^\lambda(\xi^\gamma)$, $R^{\lambda\mu}(\xi^\gamma)$, $R^{\lambda\mu\alpha}(\xi^\gamma)$, and the independent fields ξ^λ, j^λ, i^λ, respectively; to determine associated tensors $\bar{R}^\lambda(\xi^\gamma)$, $\bar{R}^{\lambda\mu}(\xi^\gamma)$, $\bar{R}^{\lambda\mu\alpha}(\xi^\gamma)$ such that:*

(1) $\bar{R}^\lambda$ is a linear function of R^λ and its first derivatives with respect

[6] Professor Chanrasekhar has called my attention to the fact that the relations between his scalars Q_1, Q_2, Q_3 and the scalars B, C, D of (3.12) and F, G, H of (3.13) are

$$\begin{aligned} Q_1 &= B + C + r\mu D = -r\mu F - H, \\ Q_2 &= -C - r^{-1}\partial(r\mu C + r^2 D)/\partial\mu = H + r^{-1}\partial(r^2F + r\mu H + G)/\partial\mu, \\ Q_3 &= -D - (r^{-1}\partial/\partial r - \mu r^{-2}\partial/\partial\mu)(r\mu C + r^2 D) \\ &= F + (r^{-1}\partial/\partial r - \mu r^{-2}\partial/\partial\mu)(r^2F + G + r\mu H). \end{aligned}$$

to ξ^λ and is homogeneous in ξ^λ, j^λ and i^λ of order zero (and similarly for $\bar{R}^{\lambda\mu}$, $\bar{R}^{\lambda\mu\alpha}$ in terms of $R^{\lambda\mu}$, $R^{\lambda\mu\alpha}$, respectively);

(2) $\bar{R}^\lambda$ has identically vanishing divergence (and similarly for $\bar{R}^{\lambda\mu}$, $\bar{R}^{\lambda\mu\alpha}$);

(3) $\bar{R}^\lambda$, $\bar{R}^{\lambda\mu}$, $\bar{R}^{\lambda\mu\alpha}$ have the structure of correlation tensors.

We shall show that *the tensors $\bar{R}^\lambda$, $\bar{R}^{\lambda\mu}$, $\bar{R}^{\lambda\mu\alpha}$ formed in accord with the following rules and only in accord with these rules satisfy the conditions* (1) *through* (3).

(4) Construct the tensors $\bar{S}_\gamma$, $\bar{S}_{\gamma}{}^{\cdot\mu}$, $\bar{S}_{\gamma}{}^{\cdot\lambda\mu}$ from the given correlation tensors R^λ, $R^{\lambda\mu}$, $R^{\lambda\mu\alpha}$ and the independent vector fields ξ^λ, j^λ, i^λ according to the laws

$$(4.1)\qquad \bar{S}_\gamma = e_{\gamma\lambda\beta}R^\lambda\xi^\beta, \qquad '\bar{S}_\gamma = e_{\gamma\lambda\beta}{}'R^\lambda j^\beta, \qquad ''\bar{S}_\gamma = e_{\gamma\lambda\beta}{}''R^\lambda i^\beta,$$

$$(4.2)\qquad \bar{S}_{\gamma}{}^{\cdot\mu} = e_{\gamma\lambda\beta}R^{\mu\lambda}\xi^\beta, \qquad '\bar{S}_{\gamma}{}^{\cdot\mu} = e_{\gamma\lambda\beta}{}'R^{\mu\lambda}j^\beta, \qquad ''\bar{S}_{\gamma}{}^{\cdot\mu} = e_{\gamma\lambda\beta}{}''R^{\mu\lambda}i^\beta,$$

$$(4.3)\qquad \bar{S}_{\gamma}{}^{\cdot\lambda\mu} = e_{\gamma\alpha\beta}R^{\lambda\mu\alpha}\xi^\beta, \qquad '\bar{S}_{\gamma}{}^{\cdot\lambda\mu} = e_{\lambda\alpha\beta}{}'R^{\lambda\mu\alpha}j^\beta, \qquad ''\bar{S}_{\gamma}{}^{\cdot\lambda\mu} = e_{\gamma\alpha\beta}{}''R^{\lambda\mu\alpha}i^\beta.$$

(5) The desired tensors, $\bar{R}^\lambda$, $\bar{R}^{\lambda\mu}$, $\bar{R}^{\lambda\mu\alpha}$, are determined by

$$(4.4)\qquad \bar{R}^\alpha = e^{\alpha\gamma\rho}\bar{S}_{\gamma,\rho}, \qquad '\bar{R}^\alpha = e^{\alpha\gamma\rho}{}'\bar{S}_{\gamma,\rho}, \qquad ''\bar{R}^\alpha = e^{\alpha\gamma\rho}{}''\bar{S}_{\gamma,\rho},$$

$$(4.5)\qquad \bar{R}^{\mu\alpha} = e^{\alpha\gamma\rho}\bar{S}_{\gamma}{}^{\cdot\mu}{}_{,\rho}, \qquad '\bar{R}^{\mu\alpha} = e^{\alpha\gamma\rho}{}'\bar{S}_{\gamma}{}^{\cdot\mu}{}_{,\rho}, \qquad ''\bar{R}^{\mu\alpha} = e^{\alpha\gamma\rho}{}''\bar{S}_{\gamma}{}^{\cdot\mu}{}_{,\rho},$$

$$(4.6)\qquad \bar{R}^{\lambda\mu\alpha} = e^{\alpha\gamma\rho}\bar{S}_{\gamma}{}^{\cdot\lambda\mu}{}_{,\rho}, \qquad '\bar{R}^{\lambda\mu\alpha} = e^{\alpha\gamma\rho}{}'\bar{S}_{\gamma}{}^{\cdot\lambda\mu}{}_{,\rho}, \qquad ''\bar{R}^{\lambda\mu\alpha} = e^{\alpha\gamma\rho}{}''\bar{S}_{\gamma}{}^{\cdot\lambda\mu}{}_{,\rho}.$$

(6) The most general tensors of the desired type are of the form

$$(4.7)\qquad \bar{\bar{R}}^\lambda = a\bar{R}^\lambda + b'\bar{R}^\lambda + c''\bar{R}^\lambda,$$

$$(4.8)\qquad \bar{\bar{R}}^{\lambda\mu} = d\bar{R}^{\lambda\mu} + e'\bar{R}^{\lambda\mu} + f''\bar{R}^{\lambda\mu},$$

$$(4.9)\qquad \bar{\bar{R}}^{\lambda\mu\alpha} = g\bar{R}^{\lambda\mu\alpha} + h'\bar{R}^{\lambda\mu\alpha} + k''\bar{R}^{\lambda\mu\alpha},$$

where $a, b, c, \cdots$ are constants. The scalars of $'\bar{R}^\lambda$, $''\bar{R}^\lambda$ will be independent scalars if the scalars of $\bar{R}^\lambda$ are less than the maximum number of independent scalars. Otherwise, the scalars of $'\bar{R}^\lambda$, $''\bar{R}^\lambda$ will be linear combinations, with variable coefficients, of the scalars of $\bar{R}^\lambda$ (and similarly for the second and third order tensors).

b. *Remarks on the Above Problem.* We consider each of the statements (1) through (6) of section (a) and discuss these statements under the corresponding numbers of this section.

(1) This first statement implies that the scalars of $\bar{R}^\lambda$ are linear func-

tions of the scalars of R^λ and of the derivatives of these scalars. This will be shown in our proof (similar results apply to $\bar{R}^{\lambda\mu}$, $\bar{R}^{\lambda\mu\alpha}$).

(2) The second statement implies that $\bar{R}^\lambda$, $\bar{R}^{\lambda\mu}$, $\bar{R}^{\lambda\mu\alpha}$ are each expressed in terms of independent scalars. In this connection, anothed problem presents itself. Do $\bar{\bar{R}}^\lambda$, $\bar{\bar{R}}^{\lambda\mu}$, $\bar{\bar{R}}^{\lambda\mu\alpha}$ contain the *maximum* number of scalars? We shall discuss this in (6).

(3) It is, of course essential that $\bar{R}^\lambda$, $\bar{R}^{\lambda\mu}$, $\bar{R}^{\lambda\mu\alpha}$ etc., shall have the structure of correlation tensors. This result will follow from the fact that R^λ, $R^{\lambda\mu}$, $R^{\lambda\mu\alpha}$ will be chosen to have such structure and the assumption (1). Again, this point must be verified in our proof.

(4) There can be at most three independent vector fields associated with three-space, independent of the type of turbulence. Of course, if the turbulence involves only one vector field, the primed and double primed tensors do not appear (see section 2). In axial symmetric turbulence, the double primed tensors do not occur, but the unprimed and primed tensors do occur. The tensors R^λ, $'R^\lambda$, $''R^\lambda$ (and similarly for the second and third order tensors) are assumed to be independent. Note, the equations (4. 1) through (4. 3) annul the non-independent scalars.

(5) It should be noted that the divergence of the tensors $\bar{R}^\lambda$, $\bar{R}^{\lambda\mu}$, $\bar{R}^{\lambda\mu\alpha}$, etc., vanish identically in virtue of (4. 4) through (4. 6).

(6) We shall show that the process of section (a) generates *all correlation tensors* which depend *linearly* upon independent scalars and their derivatives and whose *divergence vanishes* when the tensors R^λ, $'R^\lambda$, $''R^\lambda$ (and those of the second and third order), are properly chosen. Then, the $\bar{\bar{R}}^\lambda$, $\bar{\bar{R}}^{\lambda\mu}$, $\bar{\bar{R}}^{\lambda\mu\alpha}$ will contain the maximum number of independent scalars. The tensors R^λ, $'R^\lambda$, $''R^\lambda$ (and similarly for the corresponding second and third order tensors) are to be chosen as follows. Construct the tensor R^λ so that this tensor contains the maximum number of *distinct* scalars, N, not all of which are actually independent. Determine the number of conditions imposed on these scalars by the continuity relations. By subtracting the number of conditions from the maximum number of distinct scalars, one obtains *the maximum number of independent scalars, say m.* If $\bar{R}^\lambda$ contains $m_1 < m$ independent scalars, then $'\bar{R}^\lambda$ must be constructed. To do this, we introduce $'R^\lambda$. This tensor has the same structure as R^λ but depends upon N new distinct scalars. Thus, we construct a tensor $'\bar{R}^\lambda$ which depends upon m_2 (where $m_2 \leqq m - m_1$)

independent scalars. Finally, we construct $''\bar{R}^\lambda$ which depends upon m_3 (where $m_3 = m - m_1 - m_2$) independent scalars. Note, the tensor $'\bar{R}^\lambda$ may *appear* to contain more than $m - m_1$ independent scalars (see 3.12). In this case, the remaining scalars are linear combinations of the m independent scalars.

5. Proof of the results.

a. *The Tensors with Identically Vanishing Divergence which Depend Linearly upon R^λ, $R^{\lambda\mu}$, $R^{\lambda\mu\alpha}$ and Their Derivatives.* We consider the situation where two independent vector fields ξ^λ, j^λ exist. The case, in which i^λ exists, is treated in the same manner. First, consider the vector $\bar{R}^\lambda$. This vector is a general linear combination of R^λ and its derivatives with respect to ξ^λ. Hence, we may write

$$\begin{aligned}(5.1)\qquad \bar{R}^\lambda = {} & AR^\lambda + B\xi^\alpha R^\lambda{}_{,\alpha} + Cj^\alpha R^\lambda{}_{,\alpha} + D\xi^\lambda R^\alpha{}_{,\alpha} \\ & + Ej^\lambda R^\alpha{}_{,\alpha} + Fg^{\lambda\mu}\xi_\alpha R^\alpha{}_{,\mu} + Gg^{\lambda\mu}j_\alpha R^\alpha{}_{,\mu},\end{aligned}$$

where $A, B, C \cdots$ are functions of ξ^λ. We note that j^λ and ξ^λ are arbitrary. Hence, consider the transformation

$$(5.2)\qquad 'j^\lambda = j^\lambda, \qquad '\xi^\lambda = k\xi^\lambda,$$

where k is an arbitrary constant. By substituting (5.2) into (5.1), we see that $\bar{R}^\lambda$ is composed of two independent vectors:

$$\begin{aligned}(5.3)\qquad \bar{R}^\lambda &= AR^\lambda + B\xi^\alpha R^\lambda{}_{,\alpha} + D\xi^\lambda R_{,\alpha} + Fg^{\lambda\mu}\xi_\alpha R^\alpha{}_{,\mu}, \\ '\bar{R}^\lambda &= Cj^{\alpha\prime} R^\lambda{}_{,\alpha} + Ej^{\lambda\prime} R^\alpha{}_{,\alpha} + Gg^{\lambda\mu}j_\alpha{}' R^\alpha{}_{,\mu}.\end{aligned}$$

Forming the divergence of the first vector in (5.3) we obtain

$$\begin{aligned}(5.4)\qquad \bar{R}^\lambda{}_{,\lambda} = {} & (A + B + 3D + F + \xi^\lambda D_{,\lambda})R^\alpha{}_{,\alpha} + (B + D)\xi^\alpha R^\lambda{}_{,\alpha\lambda} \\ & + Fg^{\lambda\mu}\xi_\alpha R^\alpha{}_{,\mu\lambda} + R^\lambda A_{,\lambda} + \cdots,\end{aligned}$$

where the dots denote terms containing the derivatives of B, F. Since R^λ is arbitrary, we may imagine R^λ expanded in a power series in which the derivatives $R^\lambda{}_{,\alpha}$, $R^\lambda{}_{,\alpha\beta}$ are arbitrary at any given point. Hence, the condition that the divergence of $\bar{R}^\lambda$ vanish for arbitrary R^λ, ξ^λ implies that

$$(5.5)\qquad F = 0, \qquad B + D = 0, \qquad A + B + 3D + F = 0,$$

and A, B, D, F are constants. These results also follow from the fact that

the identical vanishing of (5. 4) implies the vanishing of the coefficients of all invariants of (5. 4). From (5. 5), we find

$$B = -D, \qquad A = -2D. \tag{5. 6}$$

Thus, we may represent the most general vector of the first type, $\bar{R}^\lambda$, whose divergence is identically zero and which depends linearly upon R^λ and its derivatives, by

$$\bar{R}^\lambda = -2R^\lambda - \xi^\alpha R^\lambda{}_{,\alpha} + \xi^\lambda R^\alpha{}_{,\alpha}. \tag{5. 7}$$

Similarly, one may show that the most general vector of the type $'\bar{R}^\lambda$ (see 5. 3) may be represented by

$$'\bar{R}^\lambda = j^{\alpha\prime} R^\lambda{}_{,\alpha} - j^{\lambda\prime} R^\alpha{}_{,\alpha}. \tag{5. 8}$$

By carrying out a similar analysis, one may show that the only possible second and third order tensors are

$$\bar{R}^{\mu\alpha} = -2R^{\mu\alpha} - \xi^\gamma R^{\mu\alpha}{}_{,\gamma} + \xi^\alpha R^{\mu\gamma}{}_{,\gamma}, \tag{5. 9}$$

$$'\bar{R}^{\mu\alpha} = j^{\gamma\prime} R^{\mu\alpha}{}_{,\gamma} - j^{\alpha\prime} R^{\mu\gamma}{}_{,\gamma}, \tag{5. 10}$$

$$\bar{R}^{\lambda\mu\alpha} = -2R^{\lambda\mu\alpha} - \xi^\gamma R^{\lambda\mu\alpha}{}_{,\gamma} + \xi^\alpha R^{\lambda\mu\gamma}{}_{,\gamma}, \tag{5. 11}$$

$$'\bar{R}^{\lambda\mu\alpha} = j^{\gamma\prime} R^{\lambda\mu\alpha}{}_{,\gamma} - j^{\alpha\prime} R^{\lambda\mu\gamma}{}_{,\gamma}. \tag{5. 12}$$

b. *The Equivalence of Our Construction of Tensors and the Above Construction.* It is evident that one could describe the basic tensors of the theory by equations (5. 7) through (5. 12). However, the construction used in section 4 is an equivalent construction, as will be shown. Further, this previous construction is easier to evaluate and shows why the barred tensors depend upon independent scalars. In order to show the equivalence of the barred tensors in sections 4 and 5, we carry through the computations of section 4.

Consider the equations (4. 1), (4. 4). By eliminating $\bar{S}_\gamma$, $'\bar{S}_\gamma$ we find

$$\bar{R}^\alpha = -\delta_{\lambda\beta}{}^{\alpha\rho}\partial(R^\lambda\xi^\beta)/\partial\xi^\rho, \qquad '\bar{R}^\alpha = -\delta_{\lambda\beta}{}^{\alpha\rho}\partial('R^\lambda j^\beta)/\partial\xi^\rho. \tag{5. 13}$$

Replacing $\delta_{\lambda\beta}{}^{\alpha\rho}$ by $\delta_\lambda{}^\alpha\delta_\beta{}^\rho - \delta_\beta{}^\alpha\delta_\lambda{}^\rho$ and carrying through the differentiations in (5. 13), we obtain the relations (5. 7), (5. 8). Similarly, one may show that (4. 2), (4. 5) and (4. 3), (4. 6) lead to (5. 9), (5. 10) and (5. 11), (5. 12), respectively.

c. *The Scalars of $\bar{R}^\lambda$ etc., are Linear Functions of the Scalars of R^λ etc., and Their Derivatives* (see section 4b, remark 1). In order to verify this result, we return to equation (3. 7) and the equations (3. 1), (3. 2), (3. 3)

defining the structure of R^{λ}, $R^{\lambda\mu}$, $R^{\lambda\mu\alpha}$. From these equations and the formulas (5. 7), (5. 8), etc., it is evident that the scalars of $\bar{R}^{\lambda}$ etc., are linear functions of the scalars of R^{λ} and their derivatives. A similar argument can be used if three independent vector fields ξ^{λ}, j^{λ}, i^{λ} exist.

d. *Do the $\bar{R}^{\lambda}$, etc., Have the Structure of Correlation Tensors?* The argument of (c) shows that the $\bar{R}^{\lambda}$ etc., do have the structure of correlation tensors. Note, it is essential that $\bar{R}^{\lambda}$ be a *linear homogeneous function in ξ^{λ}, j^{λ}, i^{λ} of order zero* (etc., for $\bar{R}^{\lambda\mu}$, $\bar{R}^{\lambda\mu\alpha}$, see section 4a, hypothesis 1).

6. The symmetry of $\bar{R}^{\mu\alpha}$. As we have noted, the above procedure does not assure us of the symmetry of $\bar{R}^{\mu\alpha}$. When a single independent vector field ξ^{λ} exists, then $\bar{R}^{\mu\alpha}$ is symmetric. Hence, no problem exists for this case. For the case in which two independent vector fields ξ^{λ}, j^{λ} exist, the following tensor is symmetric (if $R^{\mu\alpha}$ is symmetric) and has vanishing divergence

$$(6.1)\qquad \bar{R}^{\mu\alpha} = j^{\alpha}j^{\gamma}R^{\mu\rho}{}_{,\gamma\rho} + j^{\mu}j^{\gamma}R^{\alpha\rho}{}_{,\gamma\rho} - j^{\alpha}j^{\mu}R^{\rho\gamma}{}_{,\gamma\rho} - j^{\gamma}j^{\rho}R^{\mu\alpha}{}_{,\gamma\rho}.$$

It should be noted that the tensor $\bar{R}^{\mu\alpha}$ of (6. 1) is linear in the second derivatives of $R^{\mu\alpha}$. In fact, if we change the hypothesis (1) of section (4a) to read:

(1)′ *$\bar{R}^{\mu\alpha}$ is symmetric and is a linear function of $R^{\lambda\mu}$ and its first and second derivatives with respect to ξ^{λ},*

then present computations indicate that (6. 1) furnishes the only solution to the problem. If an additional independent vector field i^{λ} is present then an additional tensor $\bar{R}^{\lambda\mu}$ can be formed. The j^{λ} in (6. 1) is merely replaced by i^{λ}.

UNIVERSITY OF MICHIGAN.

HOMOGENEOUS DIRICHLET PROBLEM FOR INHOMOGENEOUS ULTRAHYPERBOLIC EQUATION.*

By O. G. Owens.

1. **Introduction.** There is a class [1] of 4-dimensional domains G for which the solution of the Dirichlet problem for the ultrahyperbolic equation,

$$(1.1) \qquad L[u] \equiv u_{x_1x_1} + u_{x_2x_2} - u_{y_1y_1} - u_{y_2y_2} = f(x_1, x_2, y_1, y_2),$$

is unique. Closely related uniqueness questions are discussed in a paper by F. John [2]. The existence of another such class of domains is established in the present paper. Without now specifying the differentiability assumptions on $u(x_1, x_2, y_1, y_2)$, this new class of domains can be described analytically as follows: Let $G(X)$ and $G(Y)$ be regions in X and Y space, respectively (see section 2), and let $\lambda(X)$ and $\mu(Y)$ denote the characteristic values of $G(X)$ and $G(Y)$. Then, if $|\lambda(X) - \mu(Y)|$ is never zero, the Dirichlet problem for (1.1) is unique for the cross product region $G = G(X) \times G(Y)$. Moreover, if f vanishes on G^*, and if $|\lambda(X) - \mu(Y)|$ has a positive lower bound, then a solution of (1.1) exists which vanishes on G^*. The fact that regions G exist for which $|\lambda(X) - \mu(Y)|$ is bounded away from zero is shown by means of an example.

2. **Notation.** $G(X)$ will denote a simply-connected open region in 2-dimensional cartesian (x_1, x_2)-space—more briefly X-space. $x \equiv (x_1, x_2)$ will always denote a point of $G(X)$, and $x^* \equiv (x^*_1, x^*_2)$ a point of the boundary $G^*(X)$ of $G(X)$. It will be assumed that $G^*(X)$ consists of a finite number of piecewise analytic arcs.

An arbitrary characteristic value and associated characteristic function of the region $G(X)$ will be denoted by $\lambda(X)$ and $\phi(X)$. That is, $\lambda(X)$ and $\phi(X)$ are solutions of the following equations:

$$(2.1) \qquad \Delta_x \phi(x) + \lambda(X)\phi(x) = 0, \qquad x \in G(X);$$

$$\phi(x) = 0, \qquad x \in G^*(X); \quad \int_{G(X)} \phi^2(x)\,dx = 1,$$

where $\Delta_x = \partial^2/\partial x_1^2 + \partial^2/\partial x_2^2$, $dx = dx_1 dx_2$. As in the case of $G(X)$, a

* Received October 16, 1950; revised July 31, 1951.

region $G(Y)$ is defined in 2-dimensional cartesian Y-space, together with the related characteristic values $\mu(Y)$ and associated characteristic functions $\psi(y)$. The 4-dimensional cartesian cross product region consisting of all points $(x, y) = ((x_1, x_2, y_1, y_2)$ such that $x \varepsilon G(X)$ and $y \varepsilon G(Y)$ is denoted by $G = G(X) \times G(Y)$, and its boundary by G^*.

3. Uniqueness. Our proof of the uniqueness theorem involves the application of two lemmas having to do with uniform convergence of series of characteristic functions. It is convenient to introduce some more notation for the statement of these lemmas. We write $u = u(x, y) \varepsilon C^k(X, Y)$, if u has all partial derivatives of order k, and these are continuous functions of (x, y) on $\bar{G}$, the closure of G. We write $u = u(x, y) \varepsilon C^k(X)\,(C^k(Y))$, if u has all partial derivatives with respect to $x(y)$ of order k, and these are continuous functions of (x, y) on $\bar{G}$.

Henceforth in this paper we assume that

$$(3.1) \qquad \begin{aligned} u(x^*, y) &= 0, \qquad (x^*, y) \varepsilon G^*, x^* \varepsilon G^*(X); \\ u(x, y^*) &= 0, \qquad (x, y^*) \varepsilon G^*, y^* \varepsilon G^*(Y). \end{aligned}$$

The first of our two lemmas is the following:

LEMMA 1. *If $u(x, y) \varepsilon C^4(X, Y)$, then $u(x, y)$ is equal to the uniformly, absolutely convergent series*

$$(3.2) \quad u(x, y) = \sum_{n,m=1}^{\infty} a_{nm}\phi_n(x)\psi_m(y), \qquad a_{nm} = \int_{G(X)} dx\phi_n(x) \int^{G(Y)} u(x, y)\psi_m(y)dy$$

Note. As proved, Lemma 1 is true under weaker differentiability restrictions on $u(x, y)$, namely, u, $\Delta_x u$, $\Delta_y u \varepsilon C^2(X)$ and $C^2(Y)$.

Proof. As $u(x, y) \varepsilon C^2(X)$ and satisfies (3.1), it has the series representation [3]:

$$(3.4) \qquad u(x, y) = \sum_{n=1}^{\infty} \phi_n(x) \int_{G(X)} u(x, y)\phi_n(x)\,dx.$$

As $\int_{G(X)} u(x, y)\phi_n(x)\,dx \,\varepsilon\, C^2(Y)$ and vanishes on $G^*(Y)$, (3.4) can be expressed as follows:

$$(3.5) \qquad u(x, y) = \sum_{n=1}^{\infty} \phi_n(x) \sum_{m=1}^{\infty} \psi_m(y) \int_G u(x, y)\phi_n(x)\psi_m(y)\,dxdy.$$

Because of (2.1) and (3.1), the integrals in (3.5) can be so transformed by Green's theorem that $u(x,y)$ can now be expressed as follows:

$$(3.6)\qquad u(x,y)=\sum_{n=1}^{\infty}\phi_n(x)\sum_{m=1}^{\infty}\psi_m(y)\int_G[\Delta_y\Delta_x u/(\lambda_n\lambda_m)]\phi_n(x)\psi_m(y)\,dxdy.$$

We mention that the above use of Green's theorem is all right, as the first derivatives of $\phi_n(x)$ $(\psi_m(y))$ are uniformly bounded [4] on $G(X)(G(Y))$.

The expansion (3.2) will be justified by establishing the uniform absolute convergence of the series (3.6). Our proof will be based on the inequalities of Schwarz and Bessel, and is as follows:

(3.7)

$$|u(x,y)|\leq\sum_{n=1}^{\infty}|\phi_n(x)/\lambda_n|\sum_{m=1}^{\infty}|\psi_m(y)/\mu_m|\cdot|\int_{G(Y)}dy\psi_m(y)\int_{G(X)}[\Delta_y\Delta_x u]\phi_n(x)dx|$$

$$\leq\sum_{n=1}^{\infty}|\phi_n(x)/\lambda_n|\cdot\{[\sum_{m=1}^{\infty}\psi_m^2(y)/\mu_m^2]\cdot\int_{G(Y)}dy[\int_{G(X)}[\Delta_y\Delta_x u]\phi_n(x)dx]^2\}^{\frac{1}{2}}$$

$$\leq[\sum_{n=1}^{\infty}\phi_n^2(x)/\lambda_n^2]^{\frac{1}{2}}\cdot[\sum_{m=1}^{\infty}\psi_m^2(y)/\mu_m^2]^{\frac{1}{2}}\cdot[\int_{G(Y)}dy\sum_{n=1}^{\infty}\{\int_{G(X)}[\Delta_y\Delta_x u]\phi_n(x)dx\}^2]^{\frac{1}{2}}$$

$$\leq[\sum_{n=1}^{\infty}\phi_n^2(x)/\lambda_n^2]^{\frac{1}{2}}\cdot[\sum_{m=1}^{\infty}\psi_m^2(y)/\mu_m^2]^{\frac{1}{2}}\cdot[\int_{G(Y)}dy\int_{G(X)}[\Delta_y\Delta_x x]^2dx]^{\frac{1}{2}}.$$

Hence, as $\Sigma\,\phi_n^2(x)/\lambda_n^2$ and $\Sigma\,\psi_m^2(y)/\mu_m^2$ are uniformly bounded [5], Lemma 1 is now proved.

The last lemma we need is the following:

LEMMA 2. *If $u(x,y)\in C^6(X,Y)$ and $\Delta_x u=\Delta_y u$, then*

$$(3.8)\qquad H(x,y)\equiv\sum_{n,m=1}^{\infty}\lambda_n a_{nm}\phi_n(x)\psi_m(y)=-\Delta_x u,$$

$$(3.9)\qquad I(x,y)\equiv\sum_{n,m=1}^{\infty}\mu_m a_{nm}\phi_n(x)\psi_m(y)=-\Delta_y u.$$

Note. As proved, Lemma 2 is true if $u(x,y)$ is such that $u(x,y)\in C^2(X)$ and $C^2(Y)$, $\Delta_x u\in C^2(X)$, $\Delta_x^2 u\in C^2(Y)$, $\Delta_y u\in C^2(Y)$, $\Delta_y^2 u\in C^2(X)$.

Proof. We first show that the series (3.8) is uniformly absolute convergent. Because of (2.1) and (3.3),

$$(3.10)\qquad a_{nm}=-\lambda_n^{-2}\int_{G(Y)}dy\psi_m(y)\int_{G(X)}[\Delta_x u]\phi_n(x)\,dx.$$

As $\Delta_x u(x,y)_{x=x^*}=\Delta_y u(x^*,y)=0$, (3.10) can be transformed as follows:

$$(3.11) \qquad a_{nm} = \lambda_n^{-2} \int_{G(Y)} dy \psi_m(y) \int_{G(X)} \Delta_x^2 u \phi_n(x)\, dx$$

$$a_{nm} = (-\mu_m \lambda_n^2)^{-1} \int_{G(X)} dx \phi_n(x) \int_{G(Y)} [\Delta_y \Delta_x^2 u] \psi_m(y)\, dy.$$

The last integral representation being valid since $\Delta_x^2 u(x, y)_{y=y^*} = 0$.

$H(x, y)$ can now be majorized as follows:

(3.12)

$$|H(x,y)| \leq \sum_{m=1}^{\infty} |\psi_m(y)/\mu_m| \sum_{n=1}^{\infty} |\phi_n(x)/\lambda_n| \cdot \left| \int_{G(X)} dx \phi_n(x) \int_{G(Y)} [\Delta_y \Delta_x^2 u] \psi_m(y) dy \right|$$

$$\leq \sum_{m=1}^{\infty} |\psi_m(y)/\mu_m| \cdot \{ [\sum_{n=1}^{\infty} \phi_n^2(x)/\lambda_n^2] \cdot [\int_{G(X)} dx \{ \int_{G(Y)} [\Delta_y \Delta_x^2 u] \psi_m(y) dy \}^2] \}^{\frac{1}{2}}$$

$$\leq [\sum_{n=1}^{\infty} \phi_n^2(x)/\lambda_n^2]^{\frac{1}{2}} \cdot [\sum_{m=1}^{\infty} \psi_m^2(y)/\mu_m^2]^{\frac{1}{2}} \cdot [\int_{G(X)} dx \int_{G(Y)} [\Delta_y \Delta_x^2 u]^2 dy]^{\frac{1}{2}}.$$

The above inequalities are obtained, as in (3.7), by repeated applications of the inequalities of Schwarz and Bessel. The inequality (3.12) implies that the series (3.8) has the desired convergence property.

We shall now show that $H(x, y) = -\Delta_x u(x, y)$. Let $K \equiv K(x, z)$ denote the Green's function for the region $G(X)$, and recall that the characteristic value and function are solutions of the equation

$$\phi_n(x) = \lambda_n \int_{G(X)} K(x, z) \phi_n(z)\, dz.$$

Thus, because of the uniform convergence of the series $H(x, y)$,

$$(3.13) \qquad \int_{G(X)} K(x, z) H(z, y)\, dz = \sum_{n,m=1}^{\infty} a_{nm} \psi_m(y) \lambda_n \int_{G(X)} K(x, z) \phi_n(z)\, dz$$

$$= \sum_{n,m=1}^{\infty} a_{nm} \psi_m(y) \phi_n(x) = u(x, y).$$

Moreover, as $u \in C^2(X)$, $u(x, y)$ can be represented in the form

$$(3.14) \qquad -\int_{G(X)} K(x, z) \Delta_z u(z, y)\, dz = u(x, y).$$

Consequently, because of (3.13) and (3.14), we have the following identity:

$$(3.15) \qquad V(x, y) \equiv \int_{G(X)} K(x, z) [H(z, y) + \Delta_x u(x, y)_{x=z}]\, dz \equiv 0.$$

Let $g(z, y)$ denote the bracketed expression in the integrand of (3.15). Then,

$$(3.16) \qquad V(x, y) \equiv \int_{G(X)} K(x, z) g(z, y)\, dz \equiv 0.$$

Our proof that $g \equiv 0$ will follow the usual procedure of potential theory. Let Ω denote an arbitrary area with closure in $G(X)$, and Ω^* the boundary of Ω. Represent the normal derivative with respect to Ω^* by d/dn_{Ω^*}. Then,

$$\begin{aligned} 0 &= \int_{\Omega^*} dV/dn_{\Omega^*} d\Omega^* \\ &= \int_{\Omega^*} d\Omega^* \int_{G(X)} dK/dn_{\Omega^*} g(z,y)\, dz = \int_{G(X)} dz g(z,y) \int_{\Omega^*} dK/dn_{\Omega^*} d\Omega^*. \end{aligned}$$

The interchange of the order of integration being permissible because of the summability of the normal derivative of K [4]. Now, as is well-known, the last (interior) integral is -2π if $z \subset \Omega$, and 0 if $z \subset G(x) - \Omega$. Therefore,

$$-2\pi \int_{\Omega} g(z,y)\, dz = 0.$$

As Ω has been chosen arbitrarily, and as g is continuous, we conclude that $g(x,y) \equiv 0$, that is, $-\Delta_x u(x,y) \equiv H(x,y)$. As the relation $-\Delta_y u(x,y) \equiv I(x,y)$ follows in the same way, Lemma 2 is proved.

We now consider the proof of the uniqueness result, which is stated as follows:

THEOREM 1. *Assume $u(x,y)$ is of class $C^6(X,Y)$, is a solution of the equation $\Delta_x u = \Delta_y u$, $(x,y) \in G \equiv G(X) \times G(Y)$ and assumes vanishing boundary values on G^*. Then, provided $|\lambda(X) - \mu(Y)| \neq 0$, we must have $u(x,y) \equiv 0$ on G.*

Note. Under the following weaker differentiability requirements on u, Theorem 1 is still true:

$$u,\ \Delta_x u,\ \Delta_y u \in C^2(X) \text{ and } C^2(Y);\ \Delta_x^2 u \in C^2(Y),\ \Delta_y^2 u \in C^2(X). \tag{3.17}$$

In general, functions possessing the differentiability properties (3.17) on G are said to be of class D.

Proof. By Lemma 1,

$$u(x,y) = \sum_{n,m=1}^{\infty} a_{nm} \phi_n(x) \psi_m(y), \tag{3.18}$$

and by Lemma 2,

$$0 = \Delta_x u(x,y) - \Delta_y u(x,y) = \sum_{n,m=1}^{\infty} a_{nm}(\mu_m - \lambda_n) \phi_n(x) \psi_m(y). \tag{3.19}$$

Thus, as (3.19) is uniformly absolutely convergent, and as the system $\{\phi_n(x)\psi_m(y)\}$ $(n, m = 1, 2, 3, \cdots)$ is complete on G, $a_{nm}(\mu_m - \lambda_n) = 0$.

Hence, as the characteristic values λ_n and μ_m are distinct by assumption, $a_{nm} = 0$ $(n, m = 1, 2, 3, \cdots)$, that is, $u(x, y) = 0$ for all $(x, y) \in G$.

4. Existence. The existence theorem which we establish has to do with a homogeneous boundary value problem—boundary values vanishing on G^*—for the non-homogeneous equation (1.1). The proof presupposes the vanishing of f on G^*. The restriction that $G = G(X) \times G(Y)$ is such that $|\lambda(X) - \mu(Y)|$ is uniformly bounded away from zero, is an indispensable requirement in the establishment of the existence theorem. Domains with this property are exhibited in the next section.

Before stating our theorem we introduce some new symbols.

$$D_x{}^k f(x, y)\,(D_y{}^k f(x, y))$$

will denote any partial derivative of $f(x, y)$ with respect to $x(y)$ of order k; $D^k f(x, y)$ will denote any partial derivative of $f(x, y)$ of order k. $(D^0 f(x, y) = f(x, y))$.

THEOREM 2. *Assume $G = G(X) \times G(Y)$ is such that $|\lambda(X) - \mu(Y)|$ is bounded away from zero, and that $f(x, y) \in C^{24}(X, Y)$ on G. Moreover, assume that $D^k f(x, y)$, $0 \leq k \leq 22$, vanishes on G^*. Then there is a $u(x, y)$ of class D* (see (3.17)) *which vanishes on G^*, and is a solution of* (1.1) *on G.*

Proof. Assuming the existence of such a $u(x, y)$, we first determine its series expansion. If $u \in D$ and is a solution of (1.1), then the results of Lemma 2 are still valid, provided, as we have supposed, that f vanishes on G^*. Thus

$$u(x, y) = \sum_{n,m=1}^{\infty} a_{nm}\phi_n(x)\psi_m(y), \tag{4.1}$$

$$\Delta_x u - \Delta_y u = \sum_{n,m=1}^{\infty} a_{nm}(\mu_m - \lambda_n)\phi_n(x)\psi_m(y). \tag{4.2}$$

Now, as $f \in D$, Lemma 1 yields

$$f(x, y) = \sum_{n,m=1}^{\infty} b_{nm}\phi_n(x)\psi_m(y), \tag{4.3}$$

$$b_{nm} = \int_{G(X)} dx\,\phi_n(x) \int_{G(Y)} f(x, y)\psi_m(y)\,dy. \tag{4.4}$$

Consequently, as $\{\phi_n(x)\psi_m(y)\}$ is complete on G and as (1.1) is valid, we conclude that

$$a_{nm}(\mu_m - \lambda_n) = b_{nm} \qquad (n, m = 1, 2. 3, \cdots). \tag{4.5}$$

Thus, as $(\mu_m - \lambda_n)$ is never zero,

$$u(x, y) = \sum_{n,m=1}^{\infty} b_{nm}/(\mu_m - \lambda_n)\phi_n(x)\psi_m(y). \tag{4.6}$$

Our immediate purpose is to show that the restrictions imposed upon $f(x, y)$ in the statement of the theorem are sufficient to assure that $u \varepsilon D$, and that $\Delta_x u - \Delta_y u = f(x, y)$. Such a $u(x, y)$ is unique by Theorem 1.

The essential requirements for such a study are bounds on $| D^k\phi_n(x) |$ $(0 \leq k \leq 2)$. Such bounds are readily found in the literature for domains $G(X)$ having simple closed analytic curves as boundary. Moreover, such bounds are effective in the case of our example, section 5.

In the case where $G^*(X)$ consists of a finite number of analytic arcs, A. Hammerstein [4] has found the following bounds for the characteristic functions $\phi_n(x)$:

(4.7) $| \phi_n(x) | < A\lambda_n \quad (n = 1, 2, 3, \cdots)$; (4.8) $| D'\phi_n(x) | < A\lambda_n^2$.

If $G^*(X)$ consists of a simple closed analytic curve, J. Schnauder [6] has determined bounds for the second and third partial derivatives of the solution of a general elliptic equation in terms of bounds for the solution and its first partial derivatives. The general equation considered by Schauder includes as a special case the Poisson equation: $\partial^2 v/\partial x_1^2 + \partial^2 v/\partial x_2^2 = \rho(x_1, x_2)$. Consequently, because of (4.7) and (4.8), Schauder's results enable us to conclude immediately that

(4.9) $| D^2\phi_n(x) | < A\lambda_n^3$; (4.10) $| D^3\phi_n(x) | < A\lambda_n^4$.

For the special domain $G(X)$ considered in the next section, the characteristic functions $\phi_n(x)$ are given explicitly, and the fact that they satisfy the above bounds is apparent.

It is now possible to show that $u \varepsilon D$, and we commence this investigation. In a purely formal manner, (4.6) yields the equations

$$(4.11) \qquad \begin{aligned} D_x^2 u(x, y) &= \sum_{n,m=1}^{\infty} b_{nm}/(\mu_m - \lambda_n)\psi_m(y) D_x^2\phi_n(x), \\ D_y^2 u(x, y) &= \sum_{n,m=1}^{\infty} b_{nm}/(\mu_m - \lambda_n)\phi_n(x) D_y^2\psi_m(y). \end{aligned}$$

Therefore, because of (4.10), if

$$\sum_{n,m=1}^{\infty} | b_{nm} | \mu_m\lambda_n^3 < \infty, \quad \text{and} \quad \sum_{n,m=1}^{\infty} | b_{nm} | \lambda_n\mu_m^3 < \infty, \tag{4.12}$$

then $u(x,y)$ exists and $\varepsilon\, C^2(X)$ and $C^2(Y)$. Consequences of (4.12) are the next two equations,

$$\Delta_x u(x,y) = -\sum_{n,m=1}^{\infty} \lambda_n b_{nm}/(\mu_m - \lambda_n)\phi_n(x)\psi_m(y), \tag{4.13}$$

$$\Delta_y u(x,y) = -\sum_{n,m=1}^{\infty} \mu_m b_{nm}/(\mu_m - \lambda_n)\phi_n(x)\psi_m(y), \tag{4.14}$$

which show that

$$\Delta_x u - \Delta_y u = \sum_{n,m=1}^{\infty} b_{nm}\phi_n(x)\psi_m(y) = f(x,y). \tag{4.15}$$

From (4.13) and (4.14), we conclude that both $\Delta_x u$ and $\Delta_y u$ are elements of $C^2(X)$ and $C^2(Y)$, if

$$\begin{aligned} &\sum_{n,m=1}^{\infty} | b_{nm} | \mu_m \lambda_n^4 < \infty, \qquad \sum_{n,m=1}^{\infty} | b_{nm} | \lambda_n^2 \mu_m^3 < \infty, \\ &\sum_{n,m=1}^{\infty} | b_{nm} | \lambda_n \mu_m^4 < \infty, \qquad \sum_{n,m=1}^{\infty} | b_{nm} | \mu_m^2 \lambda_n^3 < \infty. \end{aligned} \tag{4.16}$$

The inequalities (4.16) imply that

$$\Delta_x^2 u = \sum_{n,m=1}^{\infty} \lambda_n^2 b_{nm}/(\mu_m - \lambda_n)\phi_n(x)\psi_m(y), \tag{4.17}$$

$$\Delta_y^2 u = \sum_{n,m=1}^{\infty} \mu_m^2 b_{nm}/(\mu_m - \lambda_n)\phi_n(x)\psi_m(y). \tag{4.18}$$

Hence, $\Delta_x^2 u\,\varepsilon\, C^2(Y)$, and $\Delta_y^2 u\,\varepsilon\, C^2(X)$, if

$$\sum_{n,m=1}^{\infty} | b_{nm} | \mu_m^3 \lambda_3^3 < \infty. \tag{4.19}$$

Therefore, it will be sufficient to show that

$$\sum_{n,m=1}^{\infty} | b_{nm} | \mu_m^4 \lambda_m^4 < \infty, \tag{4.20}$$

in order to conclude that $u(x,y)\,\varepsilon\, D$. As $\Sigma\, 1/\lambda_n^2 < \infty$ and $\Sigma\, 1/\mu_m^2 < \infty$, the inequality (4.20) is certainly true if

$$| b_{nm} | < B/(\lambda_n^6 \mu_m^6) \qquad (n, m = 1, 2, 3, \cdots). \tag{4.21}$$

From (4.4) and (2.1) we deduce that

(4.22)

$$b_{nm} = \int_{G(Y)} dy\psi_m(y) \int_{G(X)} f(x,y)\phi_n(x)dx$$
$$= -1/\lambda_n \int_{G(Y)} dy\psi_m(y) \int_{G(X)} f(x,y)\Delta_x\phi_n(x)dx$$
$$= (-1/\lambda_n) \int_{G(Y)} dy\psi_m(y) \int_{G(X)} [\Delta_x^1 f]\phi_n(x)dx$$
$$= (-1/\lambda_n)^6 \int_{G(Y)} dy\psi_m(y) \int_{G(X)} [\Delta_x^6 f]\phi_n(x)dx.$$

In turn, (4.22) implies

$$b_{nm} = (\lambda_n\mu_m)^{-6} \int_{G(X)} dx\phi_n(x) \int_{G(Y)} [\Delta_y^6(\Delta_x^6 f)]\psi_m(y)dy. \tag{4.23}$$

An application of Schwarz's inequality to (4.23) yields the desired bound (4.21), namely,

$$|b_{nm}| \leq (\lambda_n\mu_m)^{-6}[\int_G [\Delta_y^6(\Delta_x^6 f)]^2 dxdy]^{\frac{1}{2}}.$$

This result completes the proof of Theorem 2.

5. An explicit domain G. We now propose to give an example of a region $G = G(X)\times G(Y)$ for which $|\lambda(X) - \mu(Y)|$ is bounded away from zero.

Take $G(X)$ and $G(Y)$ to be squares with sides of length a and b, respectively. More precisely, let

$$G(X) \equiv [0 \leq x_1, x_2 \leq a], \qquad G(Y) \equiv [0 \leq y_1, y_2 \leq b]. \tag{5.1}$$

Then, as is well-known, the corresponding characteristic values and functions are:

$$\begin{aligned} \phi_n(x) &\equiv (2/a)\sin(n_1\pi/a x_1)\sin(n_2\pi/a x_2), \quad \lambda(X) \equiv \pi^2(n_1^2+n_2^2)/a^2,\\ \psi_m(y) &\equiv (2/b)\sin(n_3\pi/b y_1)\sin(n_4\pi/b y_2), \quad \mu(Y) \equiv \pi^2(n_3^2+n_4^2)/b^2. \end{aligned} \tag{5.2}$$

As suggested by Mr. Louis Nirenberg and Mr. George Raney, we take

$$a = 3^{-\frac{1}{2}}R \quad \text{and} \quad b = R, \tag{5.3}$$

R being any real positive number. Then

$$|\lambda(X) - \mu(Y)| = (\pi^2/R^2)|3(n_1^2+n_2^2) - (n_3^2+n_4^2)|.$$

We shall now prove that

$$(5.4) \qquad |3(n_1^2+n_2^2)-(n_3^2+n_4^2)| \qquad (n_1, n_2, n_3, n_4 = 1, 2, 3, \cdots)$$

is never zero. A consequence of this will be $|\lambda(X)-\mu(Y)| \geq \pi^2/R^2$, a positive lower bound.

Suppose there exists a non-vacuous set of points (m_1, m_2, m_3, m_4) for which (5.4) is zero. There is then a point (m_1, m_2, m_3, m_4) of this set for which $m_3^2+m_4^2$ is a minimum. Thus, as any square is congruent either to zero or one modulo three, there are integers p_3 and p_4 such that $m_3^2 = 3^2p_3^2$ and $m_4^2 = 3^2p_4^2$. Hence, $m_1^2+m_2^2 = 3(p_3^2+p_4^2)$. The same reasoning shows the existence of integers p_1 and p_2 such that $m_1^2 = 3^2p_1^2$ and $m_2^2 = 3^2p_2^2$. Consequently, $3(p_1^2+p_2^2) = (p_3^2+p_4^2) < (m_3^2+m_4^2)$, which is a contradiction.

UNIVERSITY OF WISCONSIN.

REFERENCES.

[1] O. G. Owens, "Uniqueness of solutions of ultrahyperbolic partial differential equations," *American Journal of Mathematics*, vol. 69 (1947), pp. 184-188; especially p. 185.

[2] F. John, "On linear partial differential equations with analytic coefficients," *Communications on Pure and Applied Mathematics*, vol. II (1949), pp. 209-253.

[3] R. Courant and D. Hilbert, *Methoden der Mathematischen Physik*, vol. I (1931), p. 319.

[4] A. Hammerstein, "Über Entwicklungen gegebener Funktionen nach Eigenfunktionen von Randwertaufgaben," *Mathematische Zeitschrift*, vol. 27 (1927), pp. 269-311; especially pp. 290-294.

[5] R. Courant and D. Hilbert, *Methoden der Mathematische Physik*, vol. I (1931), p. 114.

[6] J. Schauder, "Über den Zusammenhang zwischen der Eindeutigkeit und Lösbarkeit partieller Differentialgleichungen zweiter Ordnung vom elliptischen Typus," *Mathematische Annalen*, vol. 106 (1932), pp. 661-721; especially pp. 687-688.

OSCILLATION THEOREMS.*

By Choy-Tak Taam.

1. **Introduction.** In the present paper we shall study the distribution of the complex zeros of the analytic solutions $W(z)$ of

$$d^2W/dz^2 + Q(z)W = 0, \tag{1.1}$$

where $Q(z)$ is an analytic function of the complex variable z in the domain concerned.

The method we employ in this paper is to derive a pair of simultaneous differential equations (2.7) and (2.8) which are satisfied by the modulus and argument of any solution $W(z)$ along the real axis. Since the modulus of a solution and the solution itself have exactly the same zeros, the equations (2.7) and (2.8) reflect the distribution of the zeros of the solutions on the real axis. The results so obtained can be extended to the zeros of the solutions along any analytic curve, in particular. the linear segments in the z-plane.

Using equations (2.7) and (2.8), methods will be developed to determine regions in which there are no zeros and thus we may confine them to a certain smaller known region of the z-plane. These methods could to some extent complete the methods based on the Green transform as developed by E. Hille.[1]

2. **The associated differential equations.** On the real axis, equation (1.1) has the form

$$d^2W/dx^2 + Q(x) = 0, \tag{2.1}$$

where x is the real part of the complex variable z. Let $Q(z)$ be analytic along the real axis.. In the sequel, by a solution $W(x)$ of (2.1) is always meant an analytic solution, not identically zero, unless otherwise stated.

Let

$$Q(x) = q_1 + iq_2, \qquad W(x) = u + iv, \qquad W(x) = re^{i\theta}, \tag{2.2}$$

* Received January 15, 1951; revised April 6, 1951.

The author is grateful to Professor E. Hille for some suggestions in the present paper.

[1] See E. Hille, "Oscillation theorems in the complex domain," *Transactions of the American Mathematical Society*, vol. 23 (1922), pp. 350-385.

where $q_1(x)$, $q_2(x)$, $u(x)$, $v(x)$, $r(x)$, $\theta(x)$ are real and $r(x) \geqq 0$. The derivatives $W'(x)$, $W''(x)$ with respect to x at any point where $W(x) \neq 0$ are

$$(2.3) \quad W' = (r' + ir\theta')e^{i\theta}, \qquad W'' = [\{r'' - r(\theta')^2\} + i\{r\theta'' + 2r'\theta'\}]e^{i\theta}.$$

Substituting the expressions (2.2) and (2.3) in the equation (2.1) and separating the real and imaginary parts gives

$$(2.4) \quad r'' + (q_1(x) - (\theta')^2)r = 0; \qquad (2.5) \quad r\theta'' + 2r'\theta' + q_2(x)r = 0.$$

Multiplying equation (2.5) by $r(x)$, it can be rewritten in the form

$$(2.6) \qquad (\theta' r^2)' + q_2(x)r^2 = 0.$$

The equations (2.4) and (2.6) hold at every point of the real axis where $W(x) \neq 0$. At a point where $W(x) = 0$, $r'(x)$, $r''(x)$, $\theta'(x)$ and $(\theta'(x)r^2(x))'$ do not exist. We shall, however, replace $r(x)$ and $\theta'(x)$ by functions $Y(x)$ and $\Phi(x)$ respectively (to be defined shortly) such that equations (2.4) and (2.6) will hold at every point of the real axis.

The new function $Y(x)$, called the *associated solution* of $W(x)$, is defined as follows. If $W(x)$ does not vanish on the real axis, let $Y(x) = r(x) = (u^2(x) + v^2(x))^{\frac{1}{2}}$, $-\infty < x < \infty$. If $W(x)$ vanishes at the points a_i, $a_i < a_{i+1}$, define $Y(x) = r(x)$ in $a_{2k} \leqq x \leqq a_{2k+1}$ and $Y(x) = -r(x)$ in $a_{2k-1} < x < a_{2k}$, where $k = 0, \pm 1, \pm 2, \cdots$.

The associated solution $Y(x)$ and $r(x)$ and $W(x)$ have exactly the same zeros. $Y(x)$ is obviously continuous for all x, and has continuous first and second derivatives at every point where $W(x)$ does not vanish. At every such point $Y(x)$ satisfies both equations (2.4) and (2.6).

We shall prove that $Y'(x)$ and $Y''(x)$ exist and are continuous at every zero a_i of $W(x)$. Let us consider the zero a_0. Since $W'(a_0) \neq 0$, from the definition of $Y(x)$ and the Law of the Mean, we have, as $x \to a_0 +$,

$$\begin{aligned} Y'(a_0 +) &= \lim\,[u(x)u'(x) + v(x)v'(x)][u^2(x) + v^2(x)]^{-\frac{1}{2}} \\ &= \lim\,[u'(c_1)u'(x) + v'(c_2)v'(x)][(u'(c_1))^2 + (v'(c_2))^2]^{-\frac{1}{2}} \end{aligned}$$

$= [(u'(a_0))^2 + (v'(a_0))^2]^{\frac{1}{2}} = |w'(a_0)|$, where $a_0 < c_1$, $c_2 < x$. Similarly $Y'(a_0 -) = |w'(a_0)|$. But $Y'(a_0) = \lim\,[Y(x) - Y(a_0)](x - a)^{-1}$, hence $Y'(a_0) = \lim Y'(c)$ where $c \to a_0$ and $a_0 < c < x$ or $a_0 > c > x$. Thus $Y'(x)$ exists and is continuous at a_0.

We shall soon see that $\theta'(x)$ has a limit at the zeros a_i of $W(x)$. In equation (2.4) replacing $r(x)$ by $Y(x)$, we find that $\lim Y''(x) = 0$ as $x \to a_0$. By the law of the mean,

$$Y''(a_0) = \lim_{x\to a_0}[Y'(x) - Y'(a_0)](x-a)^{-1} = \lim_{c\to a_0} Y''(c)$$

$$(a_0 < c < x \text{ or } a_0 > c > x).$$

Hence $Y''(x)$ exists and is continuous at a_0.

Between any two consecutive zeros, $a_i < x < a_{i+1}$, $W(x)$ can be represented by a continuous curve not passing through the origin in the W-plane. Then $\theta(x)$ can be so determined that it becomes a continuous function of x with continuous first derivative in $a_i < x < a_{i+1}$; namely,

$$\theta'(x) = [u(x)v'(x) - v(x)u'(x)][u^2(x) + v^2(x)]^{-1} \text{ if } a_i < x < a_{i+1}.$$

By de L'Hospital's rule it is easy to prove that

$$\lim \theta'(x) = 2^{-1}[u''(a_i)v'(a_i) - v''(a_i)u'(a_i)][(u'(a_i))^2 + (v'(a_i))^2]^{-1} = 0,$$

where $x \to a_i$. Define $\Phi(x) = \theta'(x)$ if $x \neq a_i$, and $\Phi(x) = 0$ if $x = a_i$, where $i = 0, \pm 1, \pm 2, \cdots$.

Clearly $\Phi(x)$ is continuous for all x, and from equation (2.6) it can be shown that $(\Phi(x)Y^2(x))'$ is also continuous for all x. Hence for each given solution $W(x)$ of (2.1), its associated functions $Y(x)$ and $\Phi(x)$ satisfy the simultaneous differential equations

(2.7) $\quad Y'' + (q_1 - \Phi^2)Y = 0;$ $\qquad$ (2.8) $\quad (\Phi Y^2)' + q_2 Y^2 = 0.$

3. A physical interpretation of equaltions (2.4) and (2.6). Consider a particle of unit mass whose position in a plane at time x is (r, θ), which moves under a force $q_1(x)r$ directed toward the origin and a force $q_2(x)r$ perpendicular to the radius vector but in the negative direction of θ. Equation (2.4) gives the relation of the forces acting along the radius vector, and equation (2.6) expresses the connection of the rate of change of the angular momentum to the moment of force.

4. Distribution of zeros on the real axis. In this section we study the distribution of the zeros of the solutions of (2.1) on the real axis by means of the equations (2.7) and (2.8). As in section 2, x is here a real variable.

THEOREM 4.1. *If $W(x)$ is any solution of (2.1), $q_2(x) \not\equiv 0$, and if $y(x)$ is a real-valued solution of $y'' + q_1(x)y = 0$, $y(x) \not\equiv 0$, finally if the zeros of $y(x)$ are b_i, $b_i < b_{i+1}$, $(i = 0, \pm 1, \pm 2, \cdots)$, then $W(x)$ has at most one zero in $b_i \leqq x \leqq b_{i+1}$.*

Proof. Suppose that $W(x)$ has more than one zero, and that one of them is a. From (2.8),

$$\Phi(x)Y^2(x) = -\int_a^x q_2(x)Y^2(x)\,dx,$$

because $W(x)$ and its associated solution $Y(x)$ have the same zeros. Since $q_2(x) \not\equiv 0$, $Y(x) \not\equiv 0$, then $\Phi(x) \not\equiv 0$. Now consider the equations

$$(4.1) \quad Y'' + (q_1(x) - \Phi(x))Y = 0; \qquad (4.2) \quad y'' + q_1(x)y = 0,$$

which are satisfied by $Y(x)$ and $y(x)$ respectively. By Sturm's comparison theorem,[2] we find that between any two consecutive zeros of $Y(x)$, there is at least one zero of $y(x)$. If $Y(x)$ has more than one zero in $b_i \leqq x \leqq b_{i+1}$, then $y(x)$ has at least one zero in $b_i < x < b_{i+1}$. Since $y(x)$ has no zero in $b_i < x < b_{i+1}$. then $Y(x)$ and consequently $W(x)$, has at most one zero in $b_i \leqq x \leqq b_{i+1}$.

THEOREM 4.2. *If $W(x)$ is a solution of (2.1), and if $q_2(x)$ changes sign at c_i, $c_i < c_{i+1}$ $(i = 0, \pm 1, \pm 2, \cdots)$, then $W(x)W'(x)$ has at most one zero in $c_i \leqq x \leqq c_{i+1}$.*

Proof. Suppose that $W(x)W'(x)$ has at least two zeros a_1 and a_2, $a_1 < a_2$. Since $W(a_i) = 0$ implies $Y(a_i) = 0$ and $W'(a_i) = 0$ implies $\Phi(a_i) = 0$, integrating equation (2.8) gives $\int_{a_1}^{a_2} q_2(x)Y^2(x)\,dx = 0$. Hence $q_2(x)$ must change sign at least once in $a_1 < x < a_2$. If $W(x)W'(x)$ has two zeros in $c_i \leqq x \leqq c_{i+1}$, then $q_2(x)$ must change sign at least once in $c_i < x < c_{i+1}$. Since $q_2(x)$ does not change sign in $c_i < x < c_{i+1}$, then $W(x)W'(x)$ has at most one zero in $c_i \leqq x \leqq c_{i+1}$.

5. Zero-free intervals. Suppose that $W(x)$ has a known zero, and without loss of generality, it is assumed to be at the origin. The theorems in this section will provide methods to determine intervals about a known zero in which the solution $W(x)$ does not vanish again.

THEOREM 5.1. *If $W(x)$ is a solution of (2.1), $W(0) = 0$, and if $\operatorname{Max} q_1(x) = M > 0$ in $0 \leqq x \leqq a$, then $W(x) \neq 0$ in $0 < x \leqq \operatorname{Min}(a, \pi M^{-\frac{1}{2}})$ provided that $Q(x) \not\equiv M$, and $W(x) \neq 0$ in $0 < x < \pi M^{-\frac{1}{2}}$ and $Q(x) \equiv M$ provided that $\pi M^{-\frac{1}{2}} \leqq a$ and $W(x)$ has the zero $\pi M^{-\frac{1}{2}}$.*

Proof. Let $Y(x)$, $y(x)$ and $u(x)$ be the solutions, respectively, of the following differential systems:

[2] E. L. Ince, *Differential Equations* (Dover, 1944), p. 225.

$$(5.1) \qquad Y'' + (q_1(x) - \Phi^2(x))Y = 0, \qquad Y(0) = 0;$$

$$(5.2) \qquad y'' + q_1(x)y = 0, \qquad y(0) = 0;$$

$$(5.3) \qquad u'' + Mu = 0, \qquad u(0) = 0,$$

where $Y(x)$ is the associated solution of $W(x)$, and $y(x) \not\equiv 0$, $u(x) \not\equiv 0$. Denote by b, c, and d, respectively, the smallest positive zeros of $Y(x)$, $y(x)$ and $u(x)$. If $\Phi^2(x) \not\equiv 0$ and $q_1(x) \not\equiv M$, an appeal to Sturm's comparison theorem gives $0 < d < c < b$. Hence $c = b$ implies $\Phi^2(x) \equiv 0$ and $d = c$ implies $q_1(x) \equiv M$ in $0 \leqq x \leqq a$. Since $d = \pi M^{-\frac{1}{2}}$, if $W(x)$ has a zero $\pi M^{-\frac{1}{2}}$, then $d = c = b$ and hence $\Phi^2(x) \equiv 0$, $q_1(x) \equiv M$. Consequently from (2.8), $q_2(x) \equiv 0$ and $Q(x) \equiv q_1(x) \equiv M$. The rest of the theorem is now obvious.

THEOREM 5.2. *If $W(x)$ is a solution of (2.1), $W(0) = 0$, and if $q_1(x) \leqq 0$ in $0 \leqq x \leqq a$ and $q_1(x)$ changes sign at $x = a$, finally if $\operatorname{Max} q_1(x) = M > 0$ in $a \leqq x \leqq b$, then $W(x) \neq 0$ in $0 < x \leqq \operatorname{Min}(b, a + 2^{-1}\pi M^{-\frac{1}{2}})$.*

Proof. $Y(x)$ is a solution of the differential system

$$(5.4) \qquad Y'' + (q_1(x) - \Phi^2(x))Y = 0, \qquad Y(0) = 0.$$

Define $y(x) = Cx$, $C > 0$, in $0 \leqq x \leqq a$, and $y(x) = A\cos(M^{\frac{1}{2}}x + B)$ in $a < x \leqq b$, where $Ca = A\cos(M^{\frac{1}{2}}a + B)$, $C = -A\sqrt{M}\sin(M^{\frac{1}{2}}a + B)$. It is easy to see that $y(x)$ satisfies the following differential system, except at $x = a$,

$$(5.5) \qquad y'' + f(x)y = 0, \qquad y(0) = 0,$$

where $f(x) = 0$ in $0 \leqq x \leqq a$, $f(x) = M$ in $a < x \leqq b$. Following a line of reasoning similar to that in the proof of Sturm's comparison theorem, we find that the smallest positive zero of $Y(x)$ is greater than the smallest positive zero of $y(x)$. If the latter is denoted by c, then $(M^{\frac{1}{2}}c + B) - (M^{\frac{1}{2}}a + B) > \pi/2$, hence $c > a + 2^{-1}\pi M^{-\frac{1}{2}}$.

Similar zero-free intervals to the left of the origin can be determined by using Theorems 5.1 and 5.2 after obvious modifications.

THEOREM 5.3. *If $W(x)$ is a solution of (2.1), $W(0)W'(0) = 0$, and if $q_2(x)$ keeps one sign in $a \leqq x \leqq 0$, and in $0 \leqq x \leqq b$, then $W(x)W'(x) \neq 0$ in $a \leqq x < 0$, $0 < x \leqq b$.*

Proof. This theorem follows immediately from Theorem 4.2.

6. Along analytic curves, linear segments. If p is a point on an

analytic curve[8] C, $x=f(t)$, $y=g(t)$, $a\leqq t\leqq b$, then p can be represented by $z=x+iy=f(t)+ig(t)$. Along C the differential equation (1.1) has the form

$$(6.1)\qquad d^2W/dt^2+A(t)dW/dt+B(t)W=0.$$

Let $H(t)=W(t)\exp(2^{-1}\int A(t)\,dt)$. Then equation (6.1) reduces to

$$(6.2)\qquad d^2H/dt^2+J(t)H=0.$$

Hence our results obtained on the real axis can be extended immediately to the zeros along any analytic curve.

Along the ray $\theta=\theta_0$, any point p can be represented by $z=re^{i\theta_0}$, $0\leqq\theta_0<2\pi$. Along this ray, (1.1) becomes

$$(6.3)\qquad d^2W/dr^2+Q(re^{i\theta_0})e^{i2\theta_0}W=0.$$

The real and imaginary parts of $Q(r\exp(i\theta_0))\exp(i2\theta_0)$ are, respectively, $A=q_1(r,\theta_0)\cos 2\theta_0-q_2(r,\theta_0)\sin 2\theta_0$, $B=q_1(r,\theta_0)\sin 2\theta_0+q_2(r,\theta_0)\cos 2\theta_0$. where $Q(r\exp(i\theta_0))=q_1(r,\theta_0)+iq_2(r,\theta_0)$, $q_1(r,\theta_0)$ and $q_2(r,\theta_0)$ are real.

Let $W(z)$ vanish at the origin. The theorems of section 5 provide us methods to determine linear segments on the ray $\theta=\theta_0$ in which $W(z)$ does not vanish again. Let r_0 be the positive root of the equation

$$(6.4)\qquad \pi\{\operatorname{Max} A(r,\theta_0)\}^{-\frac{1}{2}}=r_0,\text{ where } 0\leqq r\leqq r_0.$$

Then by Theorem 5.1, $W(z)\neq 0$ on the ray $\theta=\theta_0$ for $0<r<r_0$ or $0<r\leqq r_0$ according as $Q(r\exp(i\theta_0))\exp(i2\theta_0)$ is a positive constant or not. Denote this linear segment by $L_1(\theta_0)$. If $A(r,\theta_0)\leqq 0$ for all $r\geqq 0$, let $L_1(\theta_0)$ be the ray $0<r<\infty$, $\theta=\theta_0$. If $A(r,\theta_0)\leqq 0$ for $0\leqq r\leqq r_1$ and $A(r,\theta_0)$ changes sign at $r=r_1$, and if r_2 is the root of the equation

$$(6.5)\qquad 2^{-1}\pi\{\operatorname{Max} A(r,\theta_0)\}^{-\frac{1}{2}}=r_2-r_1,\text{ where } r_1\leqq r\leqq r_2,$$

then by Theorem 5.2, $W(z)\neq 0$ on the ray $\theta=\theta_0$ for $0<r\leqq r_2$. Denote this linear segment by $L_2(\theta_0)$. Again if $A(r,\theta_0)\leqq 0$ for all $r\geqq 0$, let $L_2(\theta_0)$ be the ray $0<r<\infty$, $\theta=\theta_0$. If $B(r,\theta_0)$ keeps one sign for $0<r<r_1$ on the ray $\theta=\theta_0$ and changes sign at $r=r_1$, then by Theorem 5.3 $W(z)W'(z)$ does not vanish along the segment $0<r\leqq r_1$, $\theta=\theta_0$. Denote this linear segment by $L_3(\theta_0)$. If $B(r,\theta_0)\equiv 0$, we agree that $L_3(\theta_0)$ contains no points.

7. Zero-free regions. The results of section 6 enable us to construct regions of the z-plane in which $W(z)$ does not vanish.

---

[8] See W. F. Osgoods' Lehrbuch der Funktionentheorie I, p. 702.

As θ_0 varies from 0 to 2π, each linear segment $L_i(\theta_0)$, $i = 1, 2, 3$, generates a region which we denote by R_i. In view of the discussion in section 6 and the construction of R_i, the following theorem is now obvious.

THEOREM 7.1. *If $W(z)$ is a solution of (1.1), $W(0) = 0$, then $W(z) \neq 0$ in $R_1 + R_2 + R_3$, $W(z)W'(z) \neq 0$ in R_3.*

THEOREM 7.2. *If $W(z)$ is a solution of (1.1), $W(0) = 0$, and if $\operatorname{Max} | Q(z)| = M > 0$ in $| z | \leqq a$, then $W(z) \neq 0$ in $0 < | z | \leqq \operatorname{Min}(a, \pi M^{-\frac{1}{2}})$ if $Q(z) \not\equiv$ const., $W(z) \neq 0$ in $0 < | z | < \pi M^{-\frac{1}{2}}$ and it has exactly two zeros on the circle $| z | = \pi M^{-\frac{1}{2}}$ and $Q(z) = C$, $| C | = M$, if it has a zero on this circle and $\pi M^{-\frac{1}{2}} \leqq a$.*

Proof. The differential equation (1.1) along a fixed ray reduces to equation (6.3). Since

$$M = \operatorname*{Max}_{|z| \leqq a} | Q(z) | = \operatorname*{Max}_{|z| \leqq a} | Q(re^{i\theta_0})e^{i2\theta_0} | \geqq \operatorname*{Max}_{0 \leqq r \leqq a} A(r, \theta_0),$$

where $A(r, \theta_0)$ is the real part of $Q(r \exp(i\theta_0))\exp(i2\theta_0)$, hence by Theorem 5.1, we have $W(z) \neq 0$ on the ray for $0 < r \leqq \operatorname{Min}(a, \pi M^{-\frac{1}{2}})$ provided $Q(z) \not\equiv$ constant and $\operatorname{Max} A(r, \theta_0)$ for $0 \leqq r \leqq a$ is positive. If $A(r, \theta_0) \leqq 0$ in $0 \leqq r \leqq a$, the assertion above holds in view of Theorem 5.2. Letting θ_0 vary from 0 to 2π, we have the first part of the theorem.

Let $\pi M^{-\frac{1}{2}} \leqq a$ and let $W(z)$ have a zero on the circle $| z | = \pi M^{-\frac{1}{2}}$, say $\pi M^{-\frac{1}{2}} \exp(i\theta_1)$, $\theta_1 =$ constant. Applying Theorem 5.1 to equation (6.3) gives $Q(r \exp(i\theta_1))\exp(i2\theta_1) \equiv M$. Hence $Q(z)$ is a constant on the ray $\theta = \theta_1$. But $Q(z)$ is analytic, so $Q(z) = M \exp(-2i\theta_1) = C$, $| C | = M$. The second part of the theorem is now obvious.

Along a straight line parallel to the imaginary axis, $z = x_0 + iy$, equation (1.1) reduces to

$$d^2W/dy^2 - Q(x_0 + iy)W = 0. \tag{6.6}$$

If $Q(x + iy) = q_1(x, y) + iq_2(x, y)$, $W(x + iy) = u(x, y) + iv(x, y)$ and $Y(x, y) = \pm (u^2(x, y) + v^2(x, y))^{\frac{1}{2}}$, where $q_1(x, y)$, $q_2(x, y)$, $u(x, y)$ and $v(x, y)$ are real, then $Y(x_0, y)$ satisfies

$$\partial^2Y/\partial y^2 - (q_1(x_0, y) + \Phi^2(x_0, y))Y = 0. \tag{6.7}$$

The following lemma is used to prove the next theorem.

LEMMA. *If $W(z)$ is a solution of (1.1) and is real on the real axis, and if $q_1(x_0, y) \geqq 0$ for $0 \leqq y \leqq a$ and $q_1(x_0, y)$ changes sign at $y = a$,*

finally if $\operatorname{Max}\{-q_1(x_0, y)\} = M$ *for* $a \leqq y \leqq b$, *then* $W(x_0 + iy) \neq 0$ *for* $0 < y \leqq \operatorname{Min}(b, a + 2^{-1}\pi M^{-\frac{1}{2}})$.

Proof. Using the Cauchy-Riemann conditions and condition (2), it is easy to show $Y(x_0, 0)\partial Y(x_0, 0)/\partial y = 0$. Hence either $Y(x_0, 0) = 0$ or $\partial Y(x_0, 0)/\partial y = 0$, but not both.

Case I. $Y(x_0, 0) = 0$, $\partial Y(x_0, 0)/\partial y \neq 0$. The lemma follows from a proof similar to that of Theorem 5.2.

Case II. $Y(x_0, 0) \neq 0$, $\partial Y(x_0, 0)/\partial y = 0$. Define $S(y)$ *to be* $Y(x_0, 0)$ in $0 \leqq y \leqq a$, and $Y(x_0, 0) \cos(M^{\frac{1}{2}}y + D)$ in $a < y \leqq b$, where $\cos(M^{\frac{1}{2}}a + D) = 1$, $\sin(M^{\frac{1}{2}}a + D) = 0$. Then $S(y)$ satisfies the following equation, except at $y = a$,

$$d^2S/dy^2 + f(y)S = 0, \tag{6.8}$$

where $f(y) = 0$ for $0 \leqq y \leqq a$, $f(y) = M$ for $a < y \leqq b$. On the other hand $Y(x_0, y)$ is a solution of equation (6.7). Along a line of reasoning similar to that of the proof of Sturm's comparison theorem, it is seen that the smallest positive zero of $S(y)$ is less than the smallest positive zero of $Y(x_0, y)$. The smallest positive zero of $S(y)$ is $a + 2^{-1}\pi M^{-\frac{1}{2}}$. Hence the lemma is established.

After obvious modifications, the lemma applies also to the other half of the line $z = x_0 + iy$.

Let $q_1(x_0, y) \geqq 0$ for $0 \leqq y \leqq a$ and suppose it changes sign at $y = a$. Let b be a root of $2^{-1}\pi\{\operatorname{Max} - q_1(x_0, y)\}^{-\frac{1}{2}} = b - a$, where $a \leqq y \leqq b$, and denote by $P(x_0)$ the point $z = x_0 + ib$. Using a similar method, determine a point $Q(x_0)$ on the line $z = x_0 + iy$ below the real axis. The lemma shows that $W(z)$ has no complex zero along the linear segment $P(x_0)Q(x_0)$. Letting x_0 vary from α to β, the segment $P(x_0)Q(x_0)$ generates a region which is denoted by $R_{\alpha\beta}$.

THEOREM 7.3. *If* $W(z)$ *is a solution of* (1.1) *which is real on the real axis, then* $W(z)$ *has no complex zero in* $R_{\alpha\beta}$, *where* $R_{\alpha\beta}$ *is any of the regions constructed above.*

This theorem, which follows from the construction of $R_{\alpha\beta}$, holds also if $W(z)$ is imaginary, instead of real, on the real axis, because $iW(z)$ is then real.

If $W(z)$ is real or imaginary on the imaginary axis, a similar zero-free region may be constructed.

THE UNIVERSITY OF MISSOURI.

ON THE EQUATION $a^x - b^y = 1$.*

By WM. J. LEVEQUE.

1. **Introduction.** Equations similar to that of the title were apparently first considered by S. S. Pillai [1][1] in 1931. He showed then that for any fixed set of integers a, b, c, the equation $a^x - b^y = c$ has only finitely many solutions, and that for fixed a, b the number of solutions of the inequality $0 < a^x - b^y \leqq n$ is asymptotic to

$$\frac{\log^2 n}{2 \log a \log b}$$

as $n \to \infty$. A. Herschefeld [2] showed that the equation $2^x - 3^y = c$ has at most one solution for large c, and Pillai [3] extended this to $a^x - b^y = c$ for $c > c_0 = c_0(a, b)$. We show here that the equation

$$a^x - b^y = 1 \tag{1}$$

has at most one solution unless $a = 3$, $b = 2$, when there are just two. The complete statement is contained in Theorem 6.

Since there is evidently no solution of (1) if $(a, b) > 1$, we assume throughout that a and b are relatively prime and larger than unity. Then exactly one of a, b is even; the case with a even is discussed in § 2, the case with a odd in § 3. Lemma 2 of § 3 was proved by A. Bindoni [4], but the proof given here is somewhat shorter and is included for completeness.

As application of the principal result, it is shown in § 4 that the relation $\sum_{k=1}^{n} k^3 = (\sum_{k=1}^{n} k)^2$ is the only one of its kind, i. e., that it is the only instance of the equation

$$\sum_{k=1}^{n} k^s = (\sum_{k=1}^{n} k^t)^r$$

with $r > 1$.

All letters stand for rational integers. If α belongs to $f \pmod{\beta}$ (i. e., if $z = f$ is the smallest positive solution of the congruence $\alpha^z \equiv 1 \pmod{\beta}$), we write $\alpha \to f(\beta)$. The symbol $[t_1, \cdots, t_r]$ means the l. c. m. of $t_1, \cdots, t_r$. If p is prime and $p^\alpha \mid \beta$ while $p^{\alpha+1} \nmid \beta$, we write $p^\alpha \parallel \beta$.

* Received May 8, 1950.

[1] Numbers in brackets refer to the bibliography at the end of this article.

2. The case *a* even, *b* odd.

THEOREM 1. *The equation* (1) *has at most one solution if a is even and b is odd. If* $2^\alpha \parallel a$ and $2^\beta \parallel b+1$, *the only possible solution is with* $x=\beta/\alpha$, *if* $\alpha>1$, *and with* $x=1$ *or* β *if* $\alpha=1$.

First assume that (1) holds and that at least one of α, x is larger than unity; i. e., that $\alpha x \geqq 2$. Then

$$a^x = (b-1+1)^y + 1 = (b-1)^y + y(b-1)^{y-1} + \cdots + y(b-1) + 2.$$

Put $a = 2^\alpha \cdot a_1$:

$$2^{\alpha x-1}a_1{}^x = 2^{y-1}\left(\frac{b-1}{2}\right)^y + y2^{y-2}\left(\frac{b-1}{2}\right)^{y-1} + \cdots + y\frac{b-1}{2} + 1;$$

it follows that y is odd. Putting $b = 2^\beta \cdot b_1 - 1$, we have $a^x = (2^\beta b_1 - 1)^y + 1$, so that $2^\beta \parallel a^x$, $\beta = \alpha x$, $x = \beta/\alpha$.

The proof will be complete when we show that if $\alpha = 1$ and (1) has a solution with $x=1$, it has no solution with $x>1$. Assume that there is a y_0 such that $a - b^{y_0} = 1$, and that (1) holds for some $x>1$. Then $y>y_0$ and we have

$$(b^{y_0}+1)^x - b^y = 1, \qquad b^{y_0 x} + xb^{y_0(x-1)} + \cdots + xb^{y_0} - b^y = 0,$$

whence $b \mid x$. Put $x = bx'$. By the earlier part of the proof,, the only possible solution of $(a^b)^{x'} - b^y = 1$ is with $x' = \beta/\alpha'$, where $\alpha' = \alpha b = b$. But $\beta < \alpha'$ for $b>1$, so that $\alpha' \dagger \beta$.

3. The case *a* odd, *b* even.

THEOREM 2. *If* $a \to 1(b)$ *and* $b \to 2(a)$, *then* $a = b+1$.

We have $a = jb+1$; $b^2 = ha+1$, so that $b^2 - jhb - (h+1) = 0$. The discriminant of this quadratic must be a square, so that $j^2h^2 + 4h + 4 = r^2$. Since $r > jh$ we can put $r = jh + \alpha$ with $\alpha > 0$. This gives $4h+4 = 2\alpha jh + \alpha^2$, so that α is even, and $\alpha \leqq 2$ or h is negative. Since h is positive, $\alpha = 2$, $j=1$, $a = b+1$.

Now assume that there is a solution of (1), where now a is odd and b is even. Then clearly b belongs to an even exponent (mod a), and we define s_0, t_0 by the conditions $a \to s_0(b)$, $b \to 2t_0(a)$, and define s_k, t_k recursively by the conditions

$$a^{s_0\cdots s_{k-1}} \to s_k(b^{t_0\cdots t_{k-1}}), \qquad b^{t_0\cdots t_{k-1}} \to 2t_k(a^{s_0\cdots s_{k-1}}).$$

THEOREM 3. *If* (1) *is solvable, then there is a g for which* $x_0 = s_0 \cdots s_g$,

$y_0 = t_0 \cdots t_g$ *is a solution, and for every solution* (x, y) of (1), $x_0 \mid x$ *and* $y_0 \mid y$.

Let (x, y) be a solution of (1). Then clearly $s_0 \mid x$, $t_0 \mid y$; put $x_1 = x/s_0$, $y_1 = y/t_0$. If $s_0 = t_0 = 1$, then by Theorem 2, $a - b = 1$, and the theorem holds with $g = 0$. If $s_0 t_0 > 1$, then $x_1 y_1 < xy$ and we have $(a^{s_0})^{x_1} - (b^{t_0})^{y_1} = 1$. As before, $s_1 \mid x_1$, $t_1 \mid y_1$. If $s_1 = t_1 = 1$ the theorem holds with $g = 0$; otherwise, putting $x_2 = x_1/s_1$, $y_2 = y_1/t_1$, we have $(a^{s_0 s_1})^{x_2} - (b^{t_0 t_1})^{y_2} = 1$, where $x_2 y_2 < x_1 y_1$. Continuing in this fashion we must encounter a k for which $s_k = t_k = 1$, for otherwise we would have an infinite decreasing sequence of positive integers $x_i y_i$. Then $g = k - 1$.

Two problems now confront us: to investigate the g of Theorem 3 and to show that unless a is 3 and b is 2, (x_0, y_0) is the only possible solution of (I). We need some preliminary lemmas.

LEMMA 1. *If p is an odd prime and $p \nmid ab$, each of the congruences $a \equiv b \pmod{p^k}$ and $a^{p^s} \equiv b^{p^s} \pmod{p^{k+s}}$ implies the other.*

That the first congruence implies the second is easily proved by induction on s. The reverse implication is proved for $s = 1$ by using Fermat's theorem and induction on k, and this result is used to prove the general statement by induction on s.

LEMMA 2. *If $\alpha \to t(p)$, p an odd prime, and if $p^z \parallel \alpha^t - 1$, then $\alpha \to tp^{m(n-z)}(p^n)$, where $m(x)$ denotes x for $x > 0$ and 0 for $x \leqq 0$.*

Assume the hypotheses of the theorem are satisfied. If $n \leqq z$, then $p^n \mid \alpha^t - 1$. If $p^n \mid \alpha^{t'} - 1$, then $p \mid \alpha^{t'} - 1$, so that $t' \geqq t$. This proves the theorem for $n \leqq z$.

If $n > z$, Lemma 1 shows that $\alpha^{tp^{n-z}} \equiv 1 \pmod{p^n}$. We must show that $\alpha^d \not\equiv 1 \pmod{p^n}$ if d is a proper divisor of tp^{n-z}. Let $d = t_1 p^r$ with $r \leqq n - z$, $t_1 \mid t$, and assume that $\alpha^{t_1 p^r} \equiv 1 \pmod{p^n}$. By Lemma 1, $\alpha^{t_1} \equiv 1 \pmod{p^{n-r}}$. Since $n - r \geqq z$, this contradicts the definitions of t and z, unless $r = n - z$, $t = t_1$.

LEMMA 3. *If $\alpha_i \to t_i(\beta_i)$ for $i = 1, \cdots, r$ and if the β_i are pairwise coprime, then*

$$\prod_{i=1}^{r} \alpha_i \to [t_1, \cdots, t_r] \left(\operatorname{mod} \prod_{i=1}^{r} \beta_i\right).$$

This is obvious.

We have shown that any solution (x, y) of (1) is such that $x_0 \mid x, y_0 \mid y$. Putting $a^{x_0} = a_1$, $b^{y_0} = b_1$, $x/x_0 = x'$, $y/y_0 = y'$, this means that if (x, y) is a solution of (1), then (x', y') is a solution of $(b_1 + 1)^{x'} - b_1^{y'} = 1$. Theorem 4 shows that (x , y_0) is the only solution of (1) unless $a = 3$, $b = 2$.

THEOREM 4. *The only solutions of* $3^x - 2^y = 1$ *are* $(1, 1)$ *and* $(2, 3)$. *The only solution of* $(\alpha + 1)^x - \alpha^y = 1$ *with* α *even,* $\alpha > 2$ *is* $(1, 1)$.

Let p be an odd prime dividing $\alpha + 1$, and put $\alpha + 1 = mp^\gamma$, $\gamma \geqq 1$, $p \nmid m$. If $(\alpha + 1)^x - \alpha^y = 1$, then $(mp^\gamma - 1)^y \equiv -1 \pmod{p^{\gamma x}}$. Since $\alpha \rightarrow 2(p)$ and $p^\gamma \parallel \alpha^2 - 1$, Lemma 2 shows that $\alpha \rightarrow 2p^{\gamma(x-1)}(p^{\gamma x})$. Hence $y \equiv p^{\gamma(x-1)} \pmod{2p^{\gamma(x-1)}}$, and so $y = Np^{\gamma(x-1)}$. This gives

$$(\alpha + 1)^x - \alpha^{Np^{\gamma(x-1)}} = 1.$$

But $2^{3^{x-1}} > 3^x - 1$ for $x > 2$, and for any $\alpha > 2$, $\alpha^{3^{x-1}} > (\alpha + 1)^x - 1$ for $x > 1$.

We must now investigate the g of Theorem 3.

THEOREM 5. $g = 0$ *or* 1. *In either case,* $y_0 = t_0$, $x_0 = s_0 s_1$.

Let $a = \Pi\, p_\rho^{\alpha_\rho}$; here and hereafter ρ ranges from 1 to r in all products, and $[f_\rho]$ will mean the l. c. m. of $f_1, \cdots, f_r$. We are assuming that (x_0, y_0) actually is a solution of (1), so that $b^{y_0} \equiv -1 \pmod{p_\rho}$ and the exponent to which b belongs $(\operatorname{mod} p_\rho)$ must be even, say $2\tau_\rho$. Let $p_\rho^{z_\rho} \parallel b^{2\tau_\rho} - 1$; then $p_\rho^{z_\rho} \parallel b^{\tau_\rho} + 1$. By Lemma 2,

$$b \rightarrow 2\tau_\rho p_\rho^{m(\alpha_\rho - z_\rho)} \qquad (\operatorname{mod} p_\rho^{\alpha_\rho}),$$

so that, by Lemma 3, $t_0 = [\tau_\rho p_\rho^{m(\alpha_\rho - z_\rho)}]$. Now let $b \rightarrow 2\sigma_\rho\, (p_\rho^{\alpha_\rho x_0})$, for $\rho = 1, \cdots, r$. From the way it was defined, it is clear that y_0 is the smallest exponent such that $b^{y_0} \equiv -1 \pmod{a^{x_0}}$, so that $y_0 = [\sigma_\rho]$. From Lemma 2,

$$\sigma_\rho = \tau_\rho p_\rho^{m(\alpha_\rho x_0 - z_\rho)}, \qquad \rho = 1, \cdots, r$$

so that

$$y_0 = [\tau_\rho p_\rho^{m(\alpha_\rho x_0 - z_\rho)}] = t_0\, \Pi\, p_\rho^{m(\alpha_\rho x_0 - z_\rho) - m(\alpha_\rho - z_\rho)},$$

and

(2) $$a^{x_0} - b^{t_0 \Pi p_\rho^{m(\alpha_\rho x_0 - z_\rho) - m(\alpha_\rho - z_\rho)}} = 1.$$

Since $p_\rho^{z_\rho} \mid b^{\tau_\rho}+1$, both $p_\rho^{\alpha_\rho}$ and $p_\rho^{z_\rho}$ divide $b^{t_0}+1$. Let $p_\rho^{\zeta_\rho} \parallel b^{t_0}+1$; then

$$(3) \qquad \zeta_\rho \geqq \operatorname{Max}(\alpha_\rho, z_\rho)$$

and for some κ noto divisible by any p_ρ,

$$(4) \qquad b^{t_0} = \kappa \prod p_\rho^{\zeta_\rho} - 1.$$

Thus (2) gives

$$(5) \qquad a^{x_0} = (\kappa \prod p_\rho^{\zeta_\rho} - 1)^{\prod p_\rho^{m(\alpha_\rho x_0 - z_\rho) - m(\alpha_\rho - z_\rho)}} + 1.$$

By Lemma 1, for each ρ,

$$(\kappa \prod p_\rho^{\zeta_\rho} - 1)^{p_\rho^{m(\alpha_\rho x_0 - z_\rho) - m(\alpha_\rho - z_\rho)}} \equiv -1 \,(\operatorname{mod} p_\rho^{\zeta_\rho + m(\alpha_\rho x_0 - z_\rho) - m(\alpha_\rho - z_\rho)}),$$

so that (5) gives

$$p_\rho^{\alpha_\rho x_0} \equiv 0 \,(\operatorname{mod} p_\rho^{\zeta_\rho + m(\alpha_\rho x_0 - z_\rho) - m(\alpha_\rho - z_\rho)}),$$

whence $\alpha_\rho x_0 \geqq \zeta_\rho + m(\alpha_\rho x_0 - z_\rho) - m(\alpha_\rho - z_\rho)$. But by (3) and the definition of $m(u)$,

$$\zeta_\rho + m(\alpha_\rho x_0 - z_\rho) - m(\alpha_\rho - z_\rho) \geqq \begin{cases} \alpha_\rho x_0 \text{ always} \\ \alpha_\rho x_0 + 1 \text{ if } z_\rho > \alpha_\rho x_0. \end{cases}$$

Hence

$$(6) \qquad \zeta_\rho = \operatorname{Max}(\alpha_\rho, z_\rho) \text{ and } z_\rho \leqq \alpha_\rho x_0.$$

For brevity, put $\pi = \prod p_\rho^{m(\alpha_\rho x_0 - z_\rho) - m(\alpha_\rho - z_\rho)}$ and put

$$\pi_{\rho_1} = \prod_{\substack{\rho=1 \\ \rho \neq \rho_1}}^{r} p_\rho^{m(\alpha_\rho x_0 - z_\rho) - m(\alpha_\rho - z_\rho)}.$$

We will show that unless $a = 3$, $b = 2$ the only solution of (5) is with $x_0 = s_0 s_1$, by showing that otherwise the right side of (5) is larger than the left when $\kappa = 1$, and *a fortiori* when $\kappa > 1$. We have

$$(\prod p_\rho^{\zeta_\rho} - 1)^\pi + 1 > \prod (p_\rho^{\zeta_\rho} - 1)^\pi$$

and wish to investigate the circumstances under which

$$(7) \qquad p_\rho^{\alpha_\rho x_0} \geqq (p_\rho^{\zeta_\rho} - 1)^\pi.$$

Case I. For some ρ, $\alpha_\rho = z_\rho$. By (6), $\zeta_\rho = \alpha_\rho \geqq 1$. If we choose $p_\rho = 3$, $\alpha_\rho = 1$, $x_0 = 2$; $\pi_\rho = 1$, then

$$p_\rho^{\alpha_\rho x_0} = (p_\rho^{\alpha_\rho} - 1)^{p_\rho^{\alpha_\rho x_0 - \alpha_\rho} \cdot \pi_\rho} + 1.$$

Clearly any increase in p_ρ, α_ρ, x_0 or π_ρ would make

$$p_\rho^{\alpha_\rho x_0} < (p_\rho^{\alpha_\rho} - 1)^{p_\rho^{\alpha_\rho(x_0-1)} \pi_\rho} + 1,$$

and of these four quantities only x_0 can be decreased, so that the only solution of (7) with $x_0 > 1$ is that given above, which leads to the exception $3^2 - 2^3 = 1$. But if $x_0 = 1$ then $s_0 \cdots s_g = 1$, so that $s_0 = s_1 = 1$, from which it follows that $t_1 = 1$, so that $g = 0$.

Case II. For all ρ, $z_\rho > \alpha_\rho$. Then $\zeta_\rho = z_\rho \leqq \alpha_\rho x_0$ by (6). The only circumstances under which the inequality

$$p_\rho^{\alpha_\rho x_0} < (p_\rho^{z_\rho} - 1)^{p_\rho^{\alpha_\rho x_0 - z_\rho} \cdot \pi_\rho}$$

is false are that $p_\rho = 3$ or $\alpha_\rho x_0 - z_\rho = 0$, $\pi_\rho = 1$. If the only prime factor of a is 3, it is easy to see that again $\alpha_\rho x_0 - z_\rho = 0$. From (2), $a^{s_0 \cdots s_g} - b^{t_0} = 1$. Since $t_1 = 1$, $s_2 = 1$, so that $g = 0$ or 1. This completes the proof of Theorem 5.

Combining Theorems 1-5, we summarize:

THEOREM 6. *The equation* (1) *has just the two solutions* (1, 1) *and* (2, 3) *if* $a = 3$, $b = 2$. *In all other cases it has at most one solution. If a is even and b is odd, the only possible solution is with* $x = \beta/\alpha$ *if* $2^\alpha \| a$, $2^\beta \| b + 1$ *and* $\alpha > 1$, *and with* $x = 1$ *or* β *if* $\alpha = 1$. *The only possible solution if a is odd and b is even is* $(s_0 s_1, t_0)$, *where* $a \to s_0(b)$, $b \to 2t_0(a)$, $a^{s_0} \to s_1(b^{t_0})$.

4. An application. As an immediate consequence of Theorem 6, we have

THEOREM 7. *The equation*

$$b^s + 1 = (b^t + 1)^r \tag{8}$$

has no solution with $r > 1$ *if* $b > 2$; *it has the sole solution* $r = 2$, $s = 3$, $t = 1$ *if* $b = 2$.

For putting $a = b^t + 1$, (8) gives $a^r = b^s + 1$. But the equation $a^x - b^y = 1$ has the solution $(1, t)$, and therefore has no other solution unless $b = 2$, $a = 3$.

THEOREM 8. *The only solution of the identity* $\sum_{k=1}^{n} k^s = (\sum_{k=1}^{n} k^t)^r$ *with* $r > 1$ *is* $r = 2$, $s = 3$, $t = 1$.

For by taking $n = 2$ the identity reduces to equation (8) with $b = 2$.

UNIVERSITY OF MICHIGAN.

BIBLIOGRAPHY.

[1] S. S. Pillai, "On the inequality $0 < a^x - b^y \leqq n$," *Journal of the Indian Mathematical Society*, vol. 19 (1931), pp. 1-11.

[2] A. Herschefeld, "The equation $2^x - 3^y = d$," *Bulletin of the American Mathematical Society*, vol. 42 (1936), pp. 231-234.

[3] S. S. Pillai, "On $a^x - b^y = c$," *Journal of the Indian Mathematical Society*, New Series, vol. 2 (1936), pp. 119-122.

[4] A. Bindoni, "A proposito di un articolo del Prof. Loria sui caratteri de divisibilita," *Giorn. scientifico-didattico per l'incremente degli studi Mathmatici nelle Scuole Medie*, vol. 4 (1905), pp. 87-92.

SUM FUNCTIONS IN ELEMENTARY p-ADIC ANALYSIS.*

By Gordon Overholtzer.

Introduction. I. Schur in [5] introduced as the derivate of the sequence $\{a_n\}$ with respect to the number p the sequence $\{\Delta a_n\}$, the individual term of which is given by

$$(0,1) \qquad \Delta a_n = (a_{n+1} - a_n)/p^{n+1},$$

in order to generalize Fermat's theorem of elementary number theory. Fermat's theorem states that if p is a rational prime, a a rational integer, the rational numbers Δa^{p^n} are integers. Schur proved that the terms of the first $p-1$ Schur derivates of $\{a^{p^n}\}$ are integral if $(a,p)=1$, where (a,p) is the greatest common divisor of a and p.

The author in [4] used the Schur derivates to generalize the number theoretical result that the sum of the k-th powers (k an integer) of the integers less than and prime to p^n, p a rational prime, n a positive integer, is divisible by p^{n-1} or p^n according as $p-1$ does or does not divide k. Thus the sequence

$$(0,2) \qquad \{[(i,p)=1, 1 \leqq i \leqq p^n] \sum_i i^k/p^n\},$$

has denominators at most p. (Summation conditions will frequently be written before the summation sign. If the summation conditions are not given, they remain those previously indicated for that summation index.) It was shown that the Schur derivates of all orders of the sequence (0, 2) are p-adically bounded. This result followed from a general formula for the Schur derivates of the sequence

$$(0,3) \qquad \{[(i,p)=1, 1 \leqq i \leqq p^n] \sum_i g(i)/p^n\},$$

where g is a function defined on the p-adic integers satisfying certain analyticity conditions. It was noted that the function $g(x)$ might be defined by different power series in different residue classes.

In the present paper, sums of functions defined by power series where

* Received March 17, 1950. This paper forms a part of the author's doctoral dissertation written at Indiana University. The author wishes to take this opportunity to thank Professor M. A. Zorn for his aid and encouragement in the preparation of this paper.

the argument is restricted to a single residue class module p^e are studied. It is found that the Schur derivates of the sequence

$$(0,4) \qquad \{[i \equiv i_0 (\operatorname{mod} p^e), 0 \leqq i \leqq p^n - 1] \sum_i g(i)/p^n\},$$

formed by summation in a single residue class up to p^n and division by p^n, are given by formulae analogous to those for the derivates of $(0,3)$. The sum over several residue classes may be defined additively.

A part of the theory is extended to generalized representatives assigned to the non-negative integers by a recursion using the development of the integers in the number base p, given an initial set of representatives assigned to $0, \cdots, p-1$, and an element with the same valuation as p. The generalization of arguments over which sums are to be taken is desirable for discussion of the $\mathfrak{p}$-adic analogs of the p-adic theorems given here. The $\mathfrak{p}$-adic case is not discussed in this paper.

A general expression in terms of Bernoullian numbers is found for the p-adic limit of the sequence $(0,4)$, and evaluated for certain special functions $g(x)$. The limit for $g(x) = \exp x$, $i \equiv 0 \pmod p$ is evaluated by two methods. Equality of the two results yields the same equation as that given by a power series identity on substituting $x = p$, for which both members are defined p-adically.

The limit for $g(x) = \log x$, $i \equiv 1 \pmod p$ suggests the asymptotic Stirling series for $\log \Gamma(x)$, where $\Gamma(x)$ is the classical gamma-function. If Γ_p is defined for positive rational integers by $\Gamma_p(n) = \prod_{i=0}^{n-1} (1 + ip)$, the limit of the divided sums $(0,4)$ for $g(x) = \log x$ is given by $p^{-1} \lim \log \Gamma_p(p^n)/p^n$, where $n \to \infty$. The function $\Gamma_p(n)$ is shown to be uniformly continuous and hence capable of a uniformly continuous extension to the p-adic integers. The proof depends upon a generalization of Wilson's theorem.

The examples studied suggest consideration of the difference equation

$$(0,5) \qquad f(x+1) - f(x) = g(x),$$

with a given initial condition. If g is continuous, the difference equation $(0,5)$ has exactly one continuous solution in p-adic analysis, given by the continuous extension of $S(n,g) = \sum_{i=0}^{n-1} g(i)$, which is the function formed by summing $g(x)$ over the integers up to n. If $f(x)$ is given by a power series with somewhat restricted coefficients, the sum function $S(x,f)$ is analytic, that is, given by a power series. For generalized representatives for which the initial representative set is complete, the sum function of a continuous function is continuous.

Some limit results permit the interpretation that $\lim S(p^n, g)/p^n$, where $n \to \infty$, is the derivative of $S(x, g)$ at zero.

1. Representatives, sums, and Schur derivates. In this section, the manner of assignment of representatives to the non-negative integers is described. Notation for the sums of functions $g(x)$ with the argument restricted to the representatives of a single residue class modulo p^e is introduced. The Schur derivates of all orders of the sequence formed by summation in a single residue class up to p^n of the values of a power series, are given as infinite series in sums of values of the derivatives of the power series.

Throughout the paper p will denote a rational prime, and the analysis will be understood to be p-adic. In particular, $|a|$ will be understood to be the p-adic valuation of a. Let $r_0, \cdots, r_{p-1}$ be an arbitrary set of p-adic integers. Express the positive rational integer i in the number base p. If $i = \sum_{k=0}^{e} i_k p^k$, $0 \leqq i_k \leqq p-1$, and $i \geqq p$, define $r_i = \sum_{k=0}^{e} r_{i_k} q^k$, where $|q| = |p|$. A set of p-adic integers r_i assigned to the integers according to this definition is called a set of generalized representatives. If the set of initial representatives $r_0, \cdots, r_{p-1}$, forms a complete residue set, the representatives are said to be initially complete.

If, in particular, $r_i = i$, $i = 0, \cdots, p-1$, and $q = p$, then $r_i = i$ for all integers i. These representatives will be called standard.

(1, 1) Theorem. *If a set of generalized representatives is initially complete, then $r_{i'} \equiv r_{i''} \pmod{p^e}$ if and only if $i' \equiv i'' \pmod{p^e}$.*

$$\text{Write} \quad i' = \sum_{k=0}^{e'} i_k' p^k, \; i'' = \sum_{k=0}^{e''} i_k'' p^k, \; 0 \leqq i_k', i_k'' \leqq p-1.$$

Then if $r_{i'} \equiv r_{i''} \pmod{p^e}$,

$$\sum_{k=0}^{e-1} (r_{i_k'} - r_{i_k''}) q^k + \sum_{k=e}^{e'} r_{i_k'} q^k - \sum_{k=e}^{e''} r_{i_k''} q^k \equiv 0 \pmod{p^e}.$$

Since $|q| = |p|$, and the same congruence holds to the modulus p, $r_{i_0'} \equiv r_{i_0''} \pmod{p}$, and $r_{i_0'} = r_{i_0''}$ or $i_0' = i_0''$, since $r_0, \cdots, r_{p-1}$ form a complete residue set. Similarly $i_k' = i_k''$, $k = 1, \cdots, e-1$. Then $i' \equiv i'' \pmod{p^e}$.

If $i' \equiv i'' \pmod{p^e}$, $i_k' = i_k''$, $k = 0, \cdots, e-1$. Then

$$r_{i'} - r_{i''} = \sum_{k=e}^{e'} r_{i_k'} q^k - \sum_{k=e}^{e''} r_{i_k''} q^k \equiv 0 \pmod{p^e}.$$

If the assignment of representatives to the non-negative rational integers is

regarded as a mapping from the non-negative rational integers to the p-adic integers, (1, 1) states that r is an isometry, that is, $|i'-i''|=|r_{i'}-r_{i''}|$.

Let $g(x)$ be a function defined for p-adic integers. Let us introduce an abbreviation for the sum of function values when the argument is restricted to representatives assigned to the integers in a single residue class modulo p^e and less than a given integer n. Define

$$(1,2)\qquad \sigma[i_0(p^e),n,g]=[i\equiv i_0\ (\mathrm{mod}\ p^e),\ 0\leqq i\leqq n-1]\sum_i g(r_i),$$

$$(1,3)\qquad \delta[i_0(p^e),n,g]=\sigma[i_0(p^e),n,g]/n.$$

Greek letters will indicate summation over generalized representatives; δ is a divided sum.

For summation over standard representatives, capital Roman letters will be used; thus we write

$$(1,4)\qquad S[i_0(p^e),n,g]=[i\equiv i_0\ (\mathrm{mod}\ p^e),\ 0\leqq i\leqq n-1]\sum_i g(i),$$

$$(1,5)\qquad D[i_0(p\ ),n,g]=S[i_0(p^e),n,g]/n.$$

The definitions (1, 2) through (1, 5) are for a single residue class, the neighborhood of i_0 with radius $|p^e|$. For a finite set of neighborhoods, these restricted sums may be defined by summation of the appropriate formulae.

According to (0, 1), the individual terms of the Schur derivates of $\{\delta[i_0(p^e),p^n,g]\}$ with respect to q are given by

$$\begin{aligned}\Delta^m\delta[i_0(p^e),p^n,g]&=\Delta(\Delta^{m-1}\delta[i_0(p^e),p^n,g])\\&=(\Delta^{m-1}\delta[i_0(p^e),p^{n+1},g]-\Delta^{m-1}\delta[i_0(p^e),p^n,g])/q^{n+1}.\end{aligned}$$

If $g(x)$ is given by a power series, and the divided sums $\delta[i_0(p^e),p^n,g^{(k)}]$ are bounded uniformly, the terms of the Schur derivates can be expressed as an infinite series in $\delta[i_0(p^e),p^n,g^{(k)}]$, where $g^{(k)}$ is the k-th derivative of g.

(1, 6) THEOREM. *If $g(x)$ is a function defined for $x\equiv r_{i_0}\ (\mathrm{mod}\ p^e)$ and can be expressed in a power series*

$$g(x)=\sum_{j=0}^{\infty}g^{(j)}(r_{i_0})(x-r_{i_0})^j/j!,$$

convergent if $|x-r_{i_0}|\leqq|p^e|$, $\delta[i_0(p^e),p^n,g^{(k)}]$ is bounded uniformly with respect to k, and the Schur derivates are taken with respect to q, then there exist coefficients $L^{(m)}(\gamma)$ independent of n and g such that for $n\geqq e$,

$$(1,7)\qquad \Delta^m\delta[i_0(p^e),p^n,g]=\sum_{\gamma=m}^{\infty}L^{(m)}(\gamma)q^{n(\gamma-m)}\delta[i_0(p^e),p^n,g^{(\gamma)}],$$

$$(1,8) \qquad L'(\gamma) = (pq)^{-1} \sum_{i=0}^{p-1} (r_i)^{\gamma}/\gamma\,!,$$

$$(1,9) \qquad L^{(m+1)}(\gamma) = [\mu + \nu = \gamma, \mu \geqq m, \nu \geqq 1] \sum_{\mu,\nu} q^{\mu-m} L^{(m)}(\mu) L'(\nu) + L^{(m)}(\gamma)(q^{\gamma-m-1} - q^{-1}),$$

$$(1,10) \qquad | L^{(m)}(\gamma) | \leqq | p^{-[2m+\gamma/(p-1)]} |.$$

The proof is similar to the proof of Theorem (1, 8) of [4]. To construct the proof from the one given in [4] write $\gamma[i_0(p^e), p^n, g]$ for $T[n, g(x)]$, $\delta[i_0(p^e), p^n, g]$ for $S[n, g(x)]$, r_γ for γ as an argument of g, and q for p in the power series development and the formation of the Schur derivate. The divided sums are still formed with p, which plays here the role of the number of residue classes modulo p.

The exponent of p in $L^{(m)}(\gamma) q^{n(\gamma-m)} \delta[i_0(p^e), p^n, g^{(\gamma)}]$, $\gamma \geqq m$, is at least $n(\gamma - m) - 2m - \gamma/(p-1) + N$, if $| p^N |$ is the uniform bound of $\delta[i_0(p^e), p^n, g^{(\gamma)}]$. This exponent is an increasing function of γ. Therefore all terms of the m-th Schur derivate after the first contain a power of p with exponent at least $n - 2m - (m+1)/(p-1) + N$, the value of the exponent for $\gamma = m + 1$. The exponent $n(\gamma - m) - 2m - \gamma/(p-1) + N$ is an increasing function of n. Hence $\{\Delta^m \delta[i_0(p^e), p^n, g]\}$ is bounded for all m, $\{\Delta^{m-1}\delta[i_0(p^e), p^n, g]\}$ is convergent for all m. Now

$$\lim_{n\to\infty} \Delta^m \delta[i_0(p^e), p^n, g] = \lim_{n\to\infty} \{L^{(m)}(m)\,\delta[i_0(p^e), p^n, g^{(m)}] + I p^{n-2m-(m+1)/(p-1)+N}\},$$

where $| I | \leqq 1$. Hence

$$(1,11) \qquad \lim_{n\to\infty} \Delta^m \delta[i_0(p^e), p^n, g] = L^{(m)}(m) \lim_{n\to\infty} \delta[i_0(p^e), p^n, g^{(m)}].$$

If the set of initial representatives $r_0, \cdots, r_{p-1}$ is the set consisting of zero and the $p-1$-st roots of unity, $L'(\gamma) = (pq)^{-1} \sum_{i=0}^{p-1} (r_i)^{\gamma}/\gamma\,! = 0$, if $p - 1$ does not divide γ. From the recursion (1, 9) it follows that for all m, $L^{(m)}(\gamma) = 0$, if $p - 1$ does not divide γ.

If the generalized representatives are initially complete, the requirement of uniform boundedness of $\delta[i_0(p^e), p^n, g^{(k)}]$ can be omitted from the hypothesis of Theorem (1, 6). The number-theoretical result

$$(1,12) \qquad \sum_{i=0}^{p^n-1} i^k \equiv 0 \pmod{p^{n-1}}, \ k \text{ a positive integer},$$

will be required. Since

$$\sum_{i=0}^{p^n-1} i^k = \sum_{j=0}^{n-1} p^{jk}[(\nu, p) = 1, 1 \leqq \nu \leqq p^{n-j}] \sum_{\nu} \nu^k,$$

(1, 12) follows from the fact mentioned in the introduction that

$$[(\nu, p) = 1, 1 \leqq \nu \leqq p^{n-j}] \sum_{\nu} \nu^k \equiv 0 \pmod{p^{n-j-1}}.$$

For a recent proof of this theorem, see [1]. Define t_k, k a positive integer, by $t_k(x) = x^k$. Then

$$\begin{aligned}
\sigma[i_0(p^e), p^n, t_k] &= [i \equiv i_0 \pmod{p^e}, 0 \leqq i \leqq p^n - 1] \sum_i (r_i)^k \\
&= [0 \leqq \nu \leqq p^{n-e} - 1] \sum_{\nu} (r_{i_0+\nu p^e})^k = \sum_{\nu} (r_{i_0} + r_\nu q^e)^k \\
&= \sum_{\nu} \sum_{\mu=0}^{k} C(k, \mu) (r_{i_0})^\mu \cdot (q^e)^{k-\mu} (r_\nu)^{k-\mu} \\
&= p^{n-e} r_{i_0}{}^k + \sum_{\mu=1}^{k-1} C(k, \mu) (r_{i_0})^\mu (q^e)^{k-\mu} \sum_{\nu} (r_\nu)^{k-\mu} \\
&\equiv p^{n-e} r_{i_0}{}^k + \sum_{\mu=1}^{k-1} C(k, \mu) (r_{i_0})^\mu (q^e)^{k-\mu} \sum_{\nu} (\nu)^{k-\mu} \pmod{p^{n-e}}.
\end{aligned}$$

by (1, 1), since the representatives are initially complete. An application of (1, 12) yields $\sigma[i_0(p^e), p^n, t_k] \equiv 0 \pmod{p^{n-e-1}}$, and hence $\delta[i_0(p^e), p^n, t_k] \leqq | p^{-e-1} |$. Now if g is given by the power series $\sum_{k=0}^{\infty} a_k x^k$, convergent for all p-adic integers, then $\lim_{k\to\infty} | a_k | = 0$. Let $| p^M | = \max_k | a_k |$. Then

$$| \delta[i_0(p^e), p^n, g] | = | \sum_{k=0}^{\infty} a_k \delta[i_0(p^e), p^n, t_k] | \leqq | p^M | \cdot | p^{-e-1} |.$$

Since the derivative of g is given by taking the derivative of the power series term by term, the same bound holds for the divided sums of $g^{(k)}$, for all k.

2. Evaluation of $\lim_{n\to\infty} D[i_0(p^e), p^n, g]$.

If $g(x)$ is given by a power series, the Bernoullian numbers may be introduced to evaluate $\lim_{n\to\infty} D[i_0(p^e), p^n, g]$ as an infinite series, the terms of which are products of a constant power of p, a sign factor, a Bernoullian number, and a coefficient of the power series of $g(x)$. $D[i_0(p^e), p^n, g]$ is a divided sum for the standard representatives in a residue class modulo p^e.

2, 1) THEOREM. *If $g(x)$ is a function defined for $| x - i_0 | \leqq | p^e |$ by the power series $g(i_0 + p^e y) = \sum_{k=0}^{\infty} a_k y^k$, convergent for all p-adic integers y, then*

$$(2,2)\qquad \lim_{n\to\infty} D[i_0(p^e), p^n, g] = p^{-e}[a_0 - a_1/2 + \sum_{\mu=1}^{\infty} a_{2\mu}(-1)^{\mu-1}B_\mu],$$

where B_μ is the μ-th Bernoullian number.

The power series development of $g(x)$ is used to transform $D[i_0(p^e), p^n, g]$ to a sum over all non-negative integers up to a given limit of positive powers. If $n \geqq e$,

$$D[i_0(p^e), p^n, g] = [i \equiv i_0 \pmod{p^e}, 0 \leqq i \leqq p^n - 1] \sum_i g(i)/p^n$$

$$= [0 \leqq \nu \leqq p^{n-e} - 1] \sum_\nu g(i_0 + \nu p^e)/p^n = \sum_\nu \sum_{k=0}^{\infty} a_k \nu^k/p^n$$

$$= \sum_{k=0}^{\infty} a_k \left(\sum_\nu \nu^k/p^n \right).$$

In order to justify taking a term by term limit of the right member of this equation, it is sufficient to show that

$$(2,3)\qquad \lim_{n\to\infty} a_k \sum_\nu \nu^k/p^n \text{ exists for each } k,$$

$$(2,4)\qquad \sum_{k=0}^{\infty} a_k \left(\sum_\nu \nu^k/p^n \right) \text{ converges uniformly in } n.$$

For $k = 0$, $\lim_{n\to\infty} \sum_\nu \nu^0/p^n = p^{n-e}/p^n = p^{-e}$. For $k = 1$,

$$\lim_{n\to\infty} \sum_\nu \nu/p^n = \lim_{n\to\infty} (p^{n-e} - 1)p^{n-e}/2p^n = (p^{-e}/2) \lim_{n\to\infty} (p^{n-e} - 1) = -p^{-e}/2.$$

For $k \geqq 2$, by elementary number theory,

(2, 5)

$$[0 \leqq \nu \leqq p^{n-e} - 1] \sum_\nu \nu^k/p^n = (p^{n-e})^{k+1}/(k+1) - (p^{n-e})^k/2 + [\rho \equiv 1 \pmod 2,$$

$$1 \leqq \rho \leqq k - 1] \sum_\rho (-1)^{(\rho-1)/2} B_{(\rho+1)/2} C(k, \rho)(p^{n-e})^{k-\rho}/(\rho + 1),$$

where the $C(k, \rho)$ are binomial coefficients. Hence if k is even,

(2, 6)

$$\lim_{n\to\infty} [0 \leqq \nu \leqq p^{n-e} - 1] \sum_\nu \nu^k/p^n$$

$$= \lim_{n\to\infty} (-1)^{[(k-1)-1]/2} B_{[(k-1)+1]/2} C(k, k-1)(p^{n-e})^{k-(k-1)}/p^n(k - 1 + 1)$$

$$= \lim_{n\to\infty} (-1)^{k/2-1} B_{k/2} k p^{n-e}/p^n k = (-1)^{k/2-1} p^{-e} B_{k/2},$$

and if k is odd,

$$(2,7)\qquad \lim_{n\to\infty} [0 \leqq \nu \leqq p^{n-e} - 1] \sum_\nu \nu^k/p^n = 0.$$

The limits for $k=0,1$ taken with (2, 6) and (2, 7), establish (2, 3). By (1, 12), $|a_k(\sum_\nu \nu^k/p^n)| \leqq |a_k|\,|p^{-e-1}|$, and (2, 4) follows. Since it is possible to take a term by term limit,

$$\begin{aligned}\lim_{n\to\infty} D[i_0(p^e), p^n, g] &= \sum_{k=0}^{\infty} a_k \lim_{n\to\infty} (\sum_\nu \nu^k/p^n) \\ &= p^{-e}[a_0 - a_1/2 + \sum_{\mu=1}^{\infty} a_{2\mu}(-1)^{\mu-1}B_\mu].\end{aligned}$$

This proof might have been assembled from material in [4]. (2, 6) and (2, 7) can be obtained from reexamination of (2, 8) and (2, 10) of that paper. A term-by-term limit was taken to establish (4, 2) of that paper. Combination of (4, 2), (2, 8), and (2, 10) yields

$$\begin{aligned}&\lim_{n\to\infty} [(i,p)=1, 0 \leqq i \leqq p^n - 1]\sum_i f(i)/p^n \\ &\quad = a_0(1-p^{-1}) + \sum_{j=1}^{\infty} a_{2j}(-1)^{j-1}B_j(1-p^{2j-1}), \text{ where } f(x) = \sum_{j=0}^{\infty} a_j x^j.\end{aligned}$$

The formula (2, 2) can be applied to $\exp(px) = \sum_{k=0}^{\infty} x^k(p^k/k!)$, convergent for all p-adic integers if $p \neq 2$, convergent if $x \equiv 0 \pmod 2$ for $p = 2$. (2, 2) gives

(2, 8) THEOREM. *If $p \neq 2$, then*

$$\lim_{n\to\infty} D[0(p), p^n, \exp] = p^{-1}[1 - p/2 + \sum_{\mu=1}^{\infty}(-1)^{\mu-1}B_\mu p^{2\mu}/(2\mu)!];$$

$$\lim_{n\to\infty} D[0(4), 2^n, \exp] = 2^{-2}[-1 + \sum_{\mu=1}^{\infty}(-1)^{\mu-1}B_\mu 2^{4\mu}/(2\mu)!].$$

The first limit can also be evaluated by elementary means, since

$$D[0(p), p^n, \exp] = [0 \leqq i \leqq p^{n-1} - 1]\sum_i \exp(ip)/p^n = (e^{p^n} - 1)/(e^p - 1)p^n,$$

by the formula for the sum of a geometric progression. Hence

$$(2, 9) \qquad \lim_{n\to\infty} D[0(p), p^n, \exp] = (e^p - 1)^{-1} \lim_{n\to\infty} (e^{p^n} - 1)/p^n = (e^p - 1)^{-1},$$

since $\lim_{n\to\infty} (e^{p^n} - 1)/p^n = (de^x/dx)_0 = 1$. Also

$$(2, 10) \qquad \lim_{n\to\infty} D[0(4), 2^n, \exp] = (e^4 - 1)^{-1}.$$

The formal power series identity

$$(2, 11) \qquad x/(e^x - 1) = 1 - x/2 + \sum_{\mu=1}^{\infty}(-1)^{\mu-1}B_\mu x^{2\mu}/(2\mu)!$$

is sometimes given (see [6], p. 127) as a definition of the Bernoullian numbers. Both members of (2, 11) are defined p-adically for $x = p$, p an odd prime, and for $x = 4$ when $p = 2$. Their equality follows from the two evaluations of $\lim_{n\to\infty} D[0(p), p^n, \exp]$.

For all p-adic integers α, the function $t_\alpha = (1+px)^\alpha$ is defined and given by the binomial theorem $(1+px)^\alpha = \sum_{k=0}^{\infty} C(\alpha, k)(px)^k$. An application of (2, 2) yields

(2, 12) THEOREM. *If α is a p-adic integer, then*

$$\lim_{n\to\infty} D[1(p), p^n, t_\alpha] = p^{-1}[1 - \alpha p/2 + \sum_{\mu=1}^{\infty} (-1)^{\mu-1} B_\mu p^{2\mu} C(\alpha, 2\mu)].$$

3. A p-adic γ function, Γ_p.

The function $\log x$, defined in residue class one modulo p, p odd, and in residue class one modulo 4, $p = 2$, by $\log(1+py) = \sum(-1)^{k-1}(p^k/k)y^k$ affords another application of Theorem (2, 1).

(3, 1) THEOREM. *If p is an odd rational prime, then*

$$\lim_{n\to\infty} D[1(p), p^n, \log] = p^{-1}[-p/2 + \sum_{\mu=1}^{\infty} (-1)^\mu p^{2\mu} B_\mu/2\mu];$$

$$\lim_{n\to\infty} D[1(4), 2^n, \log] = 2^{-2}[-2 + \sum_{\mu=1}^{\infty} (-1)^\mu 2^{4\mu} B_\mu/2\mu].$$

The series of (3, 1) suggest the asymptotic Stirling series for the derivative of $\log \Gamma(x)$, where $\Gamma(x)$ is the classical gamma function.

If an attempt is made to evaluate $\lim_{n\to\infty} D[1(p), p^n, \log]$, p odd, by elementary means, a product analogous to the classical Γ-function for p-adic analysis arises. By definition,

$$\begin{aligned} D[1(p), p^n, \log] &= [i \equiv 1 \pmod p, 0 \leqq i \leqq p^n - 1] \sum_i \log i/p^n \\ &= [0 \leqq \nu \leqq p^{n-1} - 1] \sum_\nu \log(1+\nu p)/p^n = \log \prod_\nu (1+\nu p)/p^n \\ &= \log \Gamma_p(p^{n-1})/p^n, \end{aligned}$$

where $\Gamma_p(n)$ is defined for positive rational integers by

$$\Gamma_p(n) = \prod_{i=0}^{n-1} (1+ip). \tag{3, 2}$$

Then

$$\lim_{n\to\infty} D[1(p), p^n, \log] = \lim_{n\to\infty} \log \Gamma_p(p^{n-1})/p^n = p^{-1} \lim_{n\to\infty} \log \Gamma_p(p^n)/p^n. \tag{3, 3}$$

For positive rational integers, $\Gamma_p(n)$ satisfies the difference equation

$$(3,4)\qquad \Gamma_p(n+1)=(np+1)\Gamma_p(n)$$

with the initial condition that $\Gamma_p(1)=1$. This difference equation is satisfied by

$$(3,5)\qquad \Gamma_p(n)=[\Gamma(1+1/p)]^{-1}p^{n-1}\Gamma(n+1/p).$$

If the asymptotic series $\log x-1/2x+\sum_{\mu=1}^{\infty}(-1)^{\mu}B_{\mu}/2\mu x^{2\mu}$ for the derivative of $\log\Gamma(x)$ (see [3:101]) is used, the formula (3,1) for p odd can be formally derived if $\lim_{n\to\infty}\log\Gamma_p(p^n)/p^n$ is interpreted as the derivative of $\log\Gamma_p(n)$ at zero.

For $p=2$, the appropriate equations are

$$\Gamma_2(n)=\prod_{i=0}^{n-1}(1+4i),\qquad \lim_{n\to\infty}D[1(4),2^n,\log]=2^{-2}\lim_{n\to\infty}\log\Gamma_2(2^n)/2^n.$$

The difference equation $\Gamma_2(n+1)=(4n+1)\Gamma_2(n)$ with the initial condition $\Gamma_2(1)=1$, is satisfied by $\Gamma_2(n)=[\Gamma(1+1/4)]^{-1}4^{n-1}\Gamma(n+1/4)$.

The function $\Gamma_p(n)$ has been defined only for rational integers. It will be shown that the definition of $\Gamma_p(n)$ can be extended to all p-adic integers.

Suppose that n_1 and n_2 are positive integers with $n_1\equiv n_2\pmod{p^e}$, $n_1\geqq n_2$. Write $n_1=n_2+kp^e$. Then

$$\Gamma_p(n_1)-\Gamma_p(n_2)=\Gamma_p(n_2)\{[n_2\leqq i\leqq n_2+kp^e-1]\prod_i(1+pi)-1\}$$

$$=\Gamma_p(n_2)\{\prod_{\rho=0}^{k-1}[n_2+\rho p^e\leqq i\leqq n_2+(\rho+1)p^e-1]\prod_i(1+pi)-1\}.$$

Since the p^e numbers $1+ip$, $i=n_2+\rho p^e,\cdots,n_2+(\rho+1)p^e-1$ are incongruent modulo p^{e+1} and are all congruent to one modulo p, they are congruent in some order to the first p^e numbers in residue class one. Therefore

$$[n_2+\rho p^e\leqq i\leqq n_2+(\rho+1)p^e-1]\prod_i(1+ip)$$
$$\equiv[0\leqq i\leqq p^e-1]\prod_i(1+ip)$$

$\pmod{p^{e+1}}$. The elements in residue class one form a subgroup of a cyclic group. Let a be a generating element. Then

$$(3,6)\qquad [0\leqq i\leqq p^e-1]\prod_i(1+ip)$$
$$\equiv[0\leqq i\leqq p^e-1]\prod_i a^i\pmod{p^{e+1}}\equiv a^{(p^e-1)p^e/2}\equiv 1$$

(mod p^{e+1}). Multiplication of the resulting congruences

$$[n_2 + \rho p^e \leqq i \leqq n_2 + (\rho+1)p^e - 1]\prod_i (1+pi) \equiv 1 \pmod{p^{e+1}},$$

$\rho = 0, \cdots, k-1$, gives the result that if $n_1 \equiv n_2 \pmod{p^e}$, $n_1 \geqq n_2$, then

$$\Gamma_p(n_1) - \Gamma_p(n_2) = \Gamma_p(n_2)\left[\prod_{i=n_2}^{n_1-1}(1+ip) - 1\right] \equiv 0 \pmod{p^{e+1}}$$

and $\Gamma_p(n_1) \equiv \Gamma_p(n_2) \pmod{p^{e+1}}$. Hence $\Gamma_p(n)$, p odd, satisfies a Lipschitz condition.

The discussion of the case $p = 2$ is similar. It suffices to show that

$$[0 \leqq i \leqq 2^e - 1]\prod_i (1+4i) \equiv 1 \pmod{2^{e+1}}.$$

The residue classes of 2^{e+2} congruent to one modulo four form a cyclic group if $e \geqq 1$ (see [2], p. 49). The number five is a generating element. Hence

$$[0 \leqq i \leqq 2^e - 1]\prod_i (1+4i) \equiv [0 \leqq i \leqq 2^e - 1]\prod_i 5^i \equiv 5^{2^e(2^e-1)/2} \equiv 1$$

(mod 2^{e+1}). Therefore $\Gamma_2(n)$ satisfies the same Lipschitz condition as $\Gamma_p(n)$, p odd. Only the weaker

(3, 6) THEOREM. $\Gamma_p(n)$ *is uniformly continuous*

will be required in the sequel. The considerations used in establishing (3, 6) are similar to the group theoretic proof of the generalized Wilson's theorem.

By a well-known topological theorem (see [7], p. 28), since $\Gamma_p(n)$ is defined and uniformly continuous on the rational integers, which are dense in the p-adic integers, Γ_p has a unique uniformly continuous extension to the p-adic integers, which will also be written as Γ_p, but with x as an argument. It will be established in the next section that $\log \Gamma_p(x)$ is analytic.

4. Continuity and analyticity of sum functions.

Let $g(x)$ be a function defined for all p-adic integers. Let us introduce as the sum function of g for generalized representatives the sum of the function values over the representatives of the integers up to n,

$$\sigma(n, g) = \sum_{i=0}^{n-1} g(r_i). \tag{4, 1}$$

For standard representatives write

$$S(n, g) = \sum_{i=0}^{n-1} g(i). \tag{4, 2}$$

Then $\sigma(n, g)$ satisfies the difference equation

$$\sigma(n+1, g) - \sigma(n, g) = g(r_n), \tag{4, 3}$$

with initial condition $\sigma(1,g)=g(r_0)$; and $S(n,g)$ is a solution of the difference equation

$$(4,4)\qquad S(n+1,g)-S(n,g)=g(n),$$

with initial condition $S(1,g)=g(0)$.

Given a particular solution of (4, 3) or (4, 4), the sum of this solution and a function of period one is again a solution. In the calculus of finite differences, a solution of (4, 4) is singled out by certain fairly complicated limit conditions (see [3], pp. 40-42). In the present discussion particular solutions are singled out by the requirement of uniform continuity. It will be shown that if $g(x)$ is uniformly continuous, the sum function $\sigma(n,g)$ is uniformly continuous.

Define the characteristic function of the neighborhood of i_0 with radius $|p^e|$ by

$$(4,5)\qquad \begin{aligned}\mathrm{Ch}[x,i_0(p^e)]&=1 \text{ if } x\equiv i_0 \pmod{p^e}\\ &=0 \text{ if } x\not\equiv i_0 \pmod{p^e}.\end{aligned}$$

Let the generalized representatives be initially complete, and let n_1 and n_2 be positive integers with $n_1\equiv n_2 \pmod{p^m}$, or $n_1=n_2+kp^m$, where k is a positive integer. If $m\geqq e$, then

$$\begin{aligned}&\sigma(n_1,\mathrm{Ch}[x,i_0(p^e)])-\sigma(n_2,\mathrm{Ch}[x,i_0(p^e)])\\ &\quad=[n_2\leqq i\leqq n_2+kp^m-1]\sum_i \mathrm{Ch}[r_i,i_0(p^e)]=kp^{m-e}\equiv 0 \pmod{p^{m-e}},\end{aligned}$$

since $r_{i_1}\equiv r_{i_2} \pmod{p^\nu}$ if and only if $i_1\equiv i_2 \pmod{p^\nu}$. This proves the

(4, 6) LEMMA. *If the generalized representatives are initially complete, then $\sigma(n,\mathrm{Ch}[i_0(p^e)])$ is uniformly continuous.*

The uniform continuity of the characteristic functions of neighborhoods will be used to prove

(4, 7) THEOREM. *If g is uniformly continous, and the generalized representatives are initially complete, then $\sigma(n,g)$ is uniformly continuous.*

Since g is uniformly continuous, for every positive integer ϵ there exists an N_1 such that if $i_1\equiv i_2 \pmod{p^{N_1}}$, so that $r_{i_1}\equiv r_{i_2} \pmod{p^{N_1}}$, then $g(r_{i_1})\equiv g(r_{i_2}) \pmod{p^\epsilon}$. Define g_ϵ by $g_\epsilon(x)=g(r_i)$ if $x\equiv r_i \pmod{p^{N_1}}$, $i=0,\cdots,p^{N_1}-1$. It follows that for all i,

$$(4,8)\qquad g(r_i)\equiv g_\epsilon(r_i) \pmod{p^\epsilon}.$$

By definition, g_ϵ is a linear combination of p^{N_1} characteristic functions; and hence, by (4, 6), $\sigma(n,g_\epsilon)$ is uniformly continuous. There exists an N_2 such

that if $n_1 \equiv n_2 \pmod{p^{N_2}}$, $n_1 \geqq n_2$, then $\sigma(n_1, g_\epsilon) \equiv \sigma(n_2, g_\epsilon) \pmod{p^\epsilon}$. But

$$\sigma(n_1, g_\epsilon) - \sigma(n_2, g_\epsilon) = \sum_{i=n_2}^{n_1-1} g_\epsilon(r_i) \equiv \sum_{i=n_2}^{n_1-1} g(r_i) \pmod{p^\epsilon} \text{ by } (4,8),$$
$$\equiv \sigma(n_1, g) - \sigma(n_2, g) \pmod{p^\epsilon}.$$

Hence if $n_1 \equiv n_2 \pmod{p^{N_2}}$, then $\sigma(n_1, g) \equiv \sigma(n_2, g) \pmod{p^\epsilon}$. Thus the sum function $\sigma(n, g)$ is uniformly continuous, and consequently is capable of a unique uniformly continuous extension $\sigma(x, g)$ to the p-adic integers. This extended function is the p-adic solution of (4, 3) singled out by the requirement of uniform continuity.

If $f(x)$ is analytic, and the coefficients of the power series satisfy certain additional restrictions, the sum function over standard representatives is analytic. Suppose that $f(x)$ is given by the power series

$$(4,9) \qquad f(x) = \sum_{i=0}^{\infty} a_i x^i,$$

convergent for all p-adic integers. The necessary and sufficient condition for convergence is that $\lim_{i \to \infty} |a_i| = 0$. Now

$$S(n, f) = \sum_{\mu=0}^{n-1} f(\mu) = \sum_{\mu=0}^{n-1} \Big(\sum_{i=0}^{\infty} a_i \mu^i \Big) = \sum_{i=0}^{\infty} a_i \Big(\sum_{\mu=0}^{n-1} \mu^i \Big),$$

and hence

(4, 10)
$$S(n, f) = n a_0 + \sum_{i=1}^{\infty} a_i \{ n^{i+1}/(i+1) - n^i/2 + \sum_{\rho=1}^{[i/2]} (-1)^{\rho-1} B_\rho C(i, 2\rho - 1) n^{i-2\rho+1}/2\rho \},$$

where $[r]$ is the greatest integer $\leqq r$.

It must be shown that (4, 10) can be rearranged to express $S(n, f)$ as a power series in n. The double series $\sum_i \sum_k b_{ik}$ can be summed in any order if the general term goes to zero, that is, if for every positive integer ϵ there exists an N such that if $i + k \geqq N$, then $|b_{ik}| \leqq |p^\epsilon|$. Let b_{ik} be the term of (4, 10) containing a_{i-1} and n^k as factors. Then the $(m+1)$-st row of the double array of b_{ik} has at most the first $m+1$ elements not zero, since the elements of the $(m+1)$-st row are the terms of the summation

$$a_m \{ n^{m+1}/(m+1) - n^m/2 + \sum_{\rho=1}^{[m/2]} (-1)^{\rho-1} B_\rho C(m, 2\rho - 1) n^{m-2\rho+1}/2\rho \}.$$

Now $|C(m, 2\rho - 1)| \leqq 1$, and $|B_\rho| \leqq |p^{-1}|$ by the von Staudt-Clausen theorem of elementary number theory. Hence for all k,

(4, 11) $\qquad | b_{m+1\, k} | \leqq | p^{-1} | \max\{| a_m/(m+1)|, \quad \max_{1 \leqq \rho \leqq [m/2]} | a_m/2\rho|\}.$

If for every positive integer ϵ there exists an N such that for all $m \geqq N$,

(4, 12) $\qquad | p^{-1} | \max\{| a_m/(m+1)|, \quad \max_{1 \leqq \rho \leqq [m/2]} | a_m/2\rho |\} \leqq | p^\epsilon |,$

then if $i + k \geqq 2N + 1$, $| b_{ik} | \leqq | p^\epsilon |$ by (4, 11) and (4, 12). If (4, 12) is satisfied by the coefficients of the power series of f, then (4, 10) can be summed in any order.

Rearrangement of (4, 10) gives

(4, 13)
$$S(n, f) = \sum_{\rho=1}^{\infty} n^\rho [a_{\rho-1}/\rho - a_\rho/2 + \sum_{k=1}^{\infty} (-1)^{k-1} a_{2k+\rho-1} B_k C(2k + \rho - 1, 2k - 1)/2k].$$

Write $S(x, f)$ for the uniformly continuous extension of $S(n, f)$ as given by the rearranged series (4, 13). Then $S(x, f)$ is given by the same power series with x as argument. The results of this discussion are summarized in

(4, 14) THEOREM. *If f is defined for all p-adic integers by the power series $f(x) = \sum_{i=0}^{\infty} a_i x^i$, and for every positive integer ϵ there exists an N such that for all $i \geqq N$,*

(4, 15) $\qquad \max\{| a_i/(i+1) |, \quad \max_{1 \leqq \rho \leqq [i/2]} | a_i/2\rho |\} \leqq | p^\epsilon |,$

then $S(x, f)$ is an analytic function given by

(4, 16)
$$S(x, f) = \sum_{\rho=1}^{\infty} x^\rho [a_{\rho-1}/\rho - a_\rho/2 + \sum_{k=1}^{\infty} (-1)^{k-1} a_{2k+\rho-1} B_k C(2k + \rho - 1, 2k - 1)/2k].$$

Note that $\lim_{n\to\infty} S(p^n, f)/p^n$ is given by the first coefficient of the power series of $S(x, f)$. This remark can be translated into a weakened statement of Theorem (2, 1). It also permits the interpretation that $\lim_{n\to\infty} S(p^n, f)/p^n$, the limit of divided sums, is the derivative of the sum function $S(x, f)$ at zero.

The analyticity of $\log \Gamma_p$ is an easy consequence of Theorem (4, 14).

(4, 17) THEOREM. *If $p \neq 2$, $\log \Gamma_p(x)$ is given by the power series*

$$\log \Gamma_p(x) = \sum_{\rho=1}^{\infty} (-1)^\rho x^\rho \{p^{\rho-1}/\rho(\rho-1) + p^\rho/2\rho + \sum_{k=1}^{\infty} (-1)^{k-1} p^{2k+\rho-1} B_k C(2k + \rho - 1, 2k - 1)/(2k + \rho - 1)2k\}.$$

Also

$$\log \Gamma_2(x) = \sum_{\rho=1}^{\infty} (-1)^{\rho} x^{\rho} \{4^{\rho-1}/\rho(\rho-1) + 4^{\rho}/2\rho$$

$$+ \sum_{k=1}^{\infty} (-1)^{k-1} 4^{2k+\rho-1} B_k C(2k+\rho-1, 2k-1)/(2k+\rho-1)2k\}.$$

For $p \neq 2$, $\log \Gamma_p(x) = S[x, \log(1+py)]$ by definition, and $\log(1+py)$ is given by the power series $\log(1+py) = \sum_{i=1}^{\infty} (-1)^{i-1}(py)^i/i$. It only remains to show that the coefficients $a_i = (-1)^{i-1}p^i/i$ satisfy the restriction (4, 15) on the coefficients of the power series. Since $|1/i| \leqq |p^{-i/p}|$, it follows that

$$\max\{|a_i/(i+1)|, \max_{1 \leqq \rho \leqq [i/2]} |a_i/2\rho|\} = \max\{|p^i/i(i+1)|, \max_{1 \leqq \rho \leqq [i/2]} |p^i/2i\rho|\}$$

$$\leqq \max\{|p^{i-(2i+1)/p}|, \max_{1 \leqq \rho \leqq [i/2]} |p^{i-(i+\rho)/p}|\} = |p^{i-(2i+1)/p}|.$$

Hence (4, 15) is satisfied for $a_i = (-1)^{i-1}p^i/i$, if $i - (2i+1)/p \geqq \epsilon$, or $i \geqq (p\epsilon + 1)/(p-2)$.

For $p = 2$; $\log \Gamma_2(x) = S[x, \log(1+4y)]$ and $a_i = (-1)^{i-1}4^i/i$. Then

$$\max\{|a_i/(i+1)|, \max_{1 \leqq \rho \leqq [i/2]} |a_i/2\rho|\} = \max\{|4^i/i(i+1)|, \max_{1 \leqq \rho \leqq [i/2]} |4^i/2i\rho|\}$$

$$\leqq \max\{|2^{2i-(i+1)/2}|, \max_{1 \leqq \rho \leqq [i/2]} |2^{2i-1+(i+\rho)/2}|\} \leqq |2^{2i-(2i+1)/2}| \leqq |2^{\epsilon}|,$$

if $2i - (2i+1)/2 \geqq \epsilon$, or $i \geqq \epsilon + 1/2$.

INDIANA UNIVERSITY.

REFERENCES.

[1] Hansraj Gupta, "An important congruence," *Proceedings of the Indian Academy of Sciences, Section A*, vol. 13 (1941), pp. 85-86.

[2] Erich Hecke, *Vorlesungen über der Theorie der algebraischen Zahlen*, Leipzig (1923).

[3] N. E. Nörlund, *Differenzenrechnung*, Berlin (1924).

[4] Gordon Overholtzer, "A new application of the Schur derivate," *Bulletin of the American Mathematical Society*, vol. 51 (1945), pp. 313-324.

[5] Issai Schur, "Ein Beitrag zur elementaren Zahlentheorie," *Sitzungberichte der Preussischen Akademie der Wissenschaften* (1933), pp. 145-151.

[6] E. T. Whittaker and G. N. Watson, *A course of modern analysis*, Cambridge (1940).

[7] G. T. Whyburn, *Analytic topology*, New York (1942).

A CLASS OF PARTIALLY ORDERED ABELIAN GROUPS RELATED TO KY FAN'S CHARACTERIZING SUBGROUPS.[1]

By A. H. Clifford.

In a recent paper,[2] Ky Fan gives three simple conditions (I, II, III) on a partially ordered abelian group which are necessary and sufficient that it be (order and group) isomorphic with a subgroup G of the group $C(\Omega)$ of all continuous real-valued functions defined on a compact[3] space Ω with two properties P_1 and P_2 (stated in **1** below). This is Ky Fan's Theorem 1 (KF, p. 418). A subgroup G of $C(\Omega)$ is called by Ky Fan a "characterizing subgroup of $C(\Omega)$" if, in addition to P_1 and P_2, it also satisfies P_3 (see **1** below; these are respectively properties (5.1), (5.3), (5.2) in KF, p. 419). The reason for the term, and the importance of the concept, is brought out by Ky Fan's Theorem 3 (KF, p. 422): if two characterizing subgroups G, G^* of $C(\Omega)$, $C(\Omega^*)$, resp., are isomorphic, then the compact spaces Ω and Ω^* are homeomorphic.

In his Theorem 4 (KF, p. 424), Ky Fan shows that a fourth condition IV is sufficient, along with I, II, and III, that G be isomorphic with a characterizing subgroup of some $C(\Omega)$ with compact Ω. But Example 1 of **3** below shows that IV is not a necessary condition. In **2** below we state a weaker condition IV′ which is necessary,[4] but Example 2 of **3** shows that IV′ is not sufficient. Incidentally, these examples are all real linear spaces; Ky Fan's paper applies almost verbatim to partially ordered spaces. The author has been unable to find a natural condition on a partially ordered abelian group G, supplementary to I, II, III, which is both necessary and sufficient that G be isomorphic with a characterizing subgroup (or subspace) of some $C(\Omega)$ with compact Ω.

In his Theorem 1, Ky Fan takes for Ω the set Λ of maximal convex

[1] Received February 12, 1951.

[2] Ky Fan, "Partially ordered additive groups of continuous functions," *Annals of Mathematics*, vol. 51 (1950), pp. 409-427. This paper will be cited as "KF." The pertinent definitions and the conditions I-IV are repeated in **1** below.

[3] In accordance with KF (p. 409), a *compact space* shall mean a Hausdorff space with the property that every covering by open sets contains a finite subcovering.

[4] The necessity of IV′ is an easy consequence of Ky Fan's Lemma 5.3 (KF, p. 420). It is also stated explicitly in Lemma 7.2 (KF, p. 423), and proved assuming IV.

subgroups of G, defining the topology on Λ by the now classical method (stated in 1) of Kakutani, Stone, and Gelfand. In his Theorem 4, he takes for Ω the set Σ of maximal singular subgroups of G, showing that Σ is a closed subset of Λ. Condition IV′ is simply that Σ be a subset of Λ. Σ is not in general closed, but, being a subspace of the compact space Λ, is completely regular. This fact is shown in **2** below (Theorem 2) without making use of the compactness of Λ. Regarded as an additive group of continuous real-valued functions on Σ, G satisfies six conditions, $P_1 - P_5$ and Q, stated in **1**. If Σ is compact, then (as Ky Fan shows) P_4, P_5, and Q are consequences of P_1, P_2, P_3, but otherwise they are completely independent thereof (examples in **3**). Groups satisfying these six conditions are characterizing of the underlying space Σ in the sense of Ky Fan's Theorem 3.

1. For the convenience of the reader, we repeat here some of Ky Fan's fundamental definitions (KF, pp. 410-411).

A *partially ordered abelian group* G is an abelian group, written additively, some of whose elements f are *non-negative* (written $f \geqq 0$) and in which (i) $0 \geqq 0$; (ii) $f \geqq 0$ and $-f \geqq 0$ imply $f = 0$; $f \geqq 0$ and $g \geqq 0$ imply $f + g \geqq 0$. $f \geqq g$ (or $g \leqq f$) means $f - g \geqq 0$. We shall denote by G^+ the set of non-negative elements of G.

An element f of G is called an *Archimedean element* of G if $f \geqq 0$ and if, for every $g \in G$, there exists a natural number n such that $nf \geqq g$. An element $f \geqq 0$ of G without this property is called a *non-Archimedean element* of G. A subgroup H of G is called a *singular subgroup* of G if (i) H contains no Archimedean element of G, and (ii) for any two elements f, g of H there is an h in H such that $h \geqq f$, $h \geqq g$. A subgroup H of G is a *convex subgroup* of G if $H \neq G$ and if $h_1 \in H$, $h_2 \in H$, $h_1 \geqq f \geqq h_2$ imply $f \in H$.

Ky Fan's conditions I, II, III, and IV are as follows.

I. *G contains a subgroup R which is isomorphic to the simply ordered additive group of all real numbers* (KF, p. 413).

II. *At least one element of R is an Archimedean element of G* (KF, p. 413).

III. *If $nf + g \geqq 0$ for all natural numbers n, then $f \geqq 0$* (KF, p. 417).

IV. *If a maximal convex subgroup N of G is not a maximal singular subgroup of G, then there is a finite set of elements $f_1, f_2, \cdots, f_n$ of G such that $f_i \notin N$ $(1 \leqq i \leqq n)$ and such that every maximal singular subgroup of G contains at least one of them* (KF, p. 423).

Let S be an abstract set, and let G be a set of real-valued functions defined on S. The Kakutani-Stone-Gelfand method of using G to define a topology in S is essentially the following. (It introduces just enough open sets in S to render the functions in G continuous.) *The neighborhoods of a point x_0 of S are subsets of S determined by x_0, any finite subset $f_1, \cdots, f_n$ of G, and any positive real number ϵ as follows: $U(x_0; f_1, \cdots, f_n; \epsilon)$ consists of all x in S satisfying $\max_{1 \leqq i \leqq n} |f_i(x) - f_i(x_0)| < \epsilon$.* We shall refer to this as the topology induced in S by G.

It will be convenient to list here a number of conditions applicable to any set G of real functions on a set S. We denote by G^+ the set of non-negative functions in G.

P_1. *G contains all constant functions.* (The unit function will be denoted by $e: e(x) = 1$ for all x in S.)

P_2. *If $x_1, x_2 \varepsilon S$ and $x_1 \neq x_2$, there exists $f \varepsilon G$ such that $f(x_1) \neq f(x_2)$.*

P_3. *If $f(x_0) = 0$ $(f \varepsilon G, x_0 \varepsilon S)$, there exists $g \varepsilon G^+$ such that $g \geqq f$ and $g(x_0) = 0$.*

P_4. *If f is a non-Archimedean element of G then $f(x_0) = 0$ for some $x_0 \varepsilon S$.*

P_5. *If a set Φ of functions in G^+ has the property that any finite subset of Φ has a common zero in S then the whole set Φ has a common zero.*

We record also the following condition, applicable if S is a topological space.

Q. *If A is a closed subset of S, and $x_0 \not\varepsilon A$, there exists a function $f \varepsilon G^+$ such that $f(x_0) = 0$ and $\inf_{y \varepsilon A} f(y) > 0$.*

THEOREM 1. *Let G be an additive group of real functions on S satisfying P_1, P_2, P_3.*

(1) *Relative to the topology of S induced by G, condition Q holds.*

(2) *For each $f \varepsilon G$ and each positive real number ϵ, the set $V(f; \epsilon)$ of all $x \varepsilon S$ such that $|f(x)| < \epsilon$ is an open subset of S. The set of all $V(f; \epsilon)$ with $f \varepsilon G^+$ and $\epsilon > 0$ is a basis for the open sets of S.*

(3) *If S is a topological space to begin with, the topology of S induced by G will coincide with its original topology if and only if, relative to the latter, the functions in G are continuous and condition Q holds.*

(4) *S is completely regular.*[5]

Proof. (1). Let A be a closed set in the topology of S induced by G, and let $x_0 \notin A$. Then there exists a neighborhood $U(x_0; f_1, \cdots, f_n; \epsilon)$ of x_0 having void intersection with A. Let $f_i(x_0) = \alpha_i$. By P_1, G contains $\alpha_i e$. Since G is an additive group, it contains the functions $f_i - \alpha_i e$, which vanish at x_0. By P_3, G^+ contains functions g_i vanishing at x_0 such that $g_i \geqq f_i - \alpha_i e$. Then the functions $h_i = g_i - f_i + \alpha_i e$ also belong to G^+. Let $f = g_1 + \cdots + g_n + h_1 + \cdots + h_n$. Then $f \varepsilon G^+$ and $f(x_0) = 0$. For any point y in A we have $y \notin U$, and hence $| f_i(y) - \alpha_i | \geqq \epsilon$ for some i.

$$f(y) \geqq g_i(y) + h_i(y) \geqq | g_i(y) - h_i(y) | = | f_i(y) - \alpha_i | \geqq \epsilon.$$

Hence $\inf_{y \varepsilon A} f(y) \geqq \epsilon > 0$.

(2). That $V(f; \epsilon)$ is open will follow when we show that to each $x_0 \varepsilon V$ there exists a neighborhood U of x_0 contained in V. Let $f(x_0) = \alpha$. Then $\eta = \epsilon - | \alpha | > 0$. As in (1), $g = f - \alpha e$ belongs to G. Let $U = U(x_0; g; \eta)$. If $x \varepsilon U$ then $| g(x) | < \eta$ and

$$| f(x) | = | g(x) + \alpha | \leqq | g(x) | + | \alpha | < \eta + | \alpha | = \epsilon$$

whence $x \varepsilon V$. To show that the sets $V(f; \epsilon)$ with $f \varepsilon G^+$ and $\epsilon > 0$ constitute a basis, we must show that if

$$x_0 \varepsilon W = V(f_1; \epsilon_1) \cap V(f_2; \epsilon_2) \qquad (f_1, f_2 \varepsilon G; \epsilon_1, \epsilon_2 > 0)$$

then there exist $f_3 \varepsilon G^+$ and $\epsilon_3 > 0$ such that $x_0 \varepsilon V(f_3; \epsilon_3) \subseteq W$. Let A be the complement of W in S. Since A is closed and $x_0 \notin A$, there exist by (1) $f_3 \varepsilon G^+$ and $\epsilon_3 > 0$ such that $\inf_{y \varepsilon A} f_3(y) \geqq \epsilon_3$, while $f(x_0) = 0$. If $x \varepsilon V(f_3; \epsilon_3)$ then $f_3(x) < \epsilon_3$, whence $x \notin A$ and $V(f_3; \epsilon_3) \subseteq W$.

(3). Assume that the functions in G are continuous and that Q holds in the original topology of S. Let N be an open subset of S in the original topology, and let $x_0 \varepsilon N$. Since Q holds, there exists $f \varepsilon G^+$ such that $f(x_0) = 0$, $f(x) \geqq \epsilon > 0$ for all x in the complement of N. Then $V(f; \epsilon) \subseteq N$, and N is open in the topology of S induced by G.

Conversely, we are to show that every subset of S open in the latter

[5] Part (4) is a consequence of two theorems in N. Bourbaki, *Topologie Générale* (Actualitées Scientifiques et Industrielles, nos. 858 and 1045). As a special case of a result on p. 117 of Chapter II, the space S is "uniformisable," while Theorem 2 on p. 9 of Chapter IX states that a space is uniformisable if and only if it is completely regular. The author is indebted to Professor Fan both for pointing out this reference and for supplying the quick proof of (4) which we give.

topology is also open in the original. By (2) it suffices to show this for every $V(f;\epsilon)$ with $f \varepsilon G^+$, $\epsilon > 0$. Let $x_0 \varepsilon V(f;\epsilon)$, i. e. $|f(x_0)| < \epsilon$. Let $\eta = \epsilon - |f(x_0)|$. Since f is continuous, by hypothesis, there exists a neighborhood N of x_0 such that $|f(x) - f(x_0)| < \eta$ if $x \varepsilon N$. Hence $x \varepsilon N$ implies

$$|f(x)| \leqq |f(x_0)| + |f(x) - f(x_0)| < (\epsilon - \eta) + \eta = \epsilon,$$

i. e. $x \varepsilon V(f;\epsilon)$, and hence $N \subseteq V(f;\epsilon)$.

(4). Let $U(x_0; f_1, \cdots, f_n; \epsilon)$ be a neighborhood of a point x_0 of S. We are to show the existence of a real continuous function g on S such that $g(x_0) = 0$, $g(x) = 1$ for $x \notin U$, and $0 \leqq g(x) \leqq 1$ for every $x \varepsilon S$. Such a function is the following:

$$g(x) = \min\{1, 1/\epsilon \max_{1 \leqq i \leqq n} |f_i(x) - f_i(x_0)|\}.$$

P_2 implies that every single-element subset of S is closed.

2. We now apply Theorem 1 to a partially ordered abelian group G satisfying Ky Fan's conditions I, II, III, and also:

IV'. *Every maximal singular subgroup of G is maximal convex.*

As in the proof of Ky Fan's Theorem 4 (KF, p. 424), G can be isomorphically represented as an additive group of real-valued functions on the set Σ of maximal singular subgroups of G. The function corresponding to $f \varepsilon G$ is denoted by $\hat{f}$. For $M \varepsilon \Sigma$, $\hat{f}(M)$ is the real number corresponding to the coset $f + M$ in the isomorphism between G/M and R.

Properties P_1 and P_2 are immediate for $S = \Sigma$, and P_3 is clear since each $M_0 \varepsilon \Sigma$ is singular, and $\hat{f}(M_0) = 0$ is equivalent to $f \varepsilon M_0$. We may now topologize Σ as in Theorem 1. Σ becomes thereby a completely regular space; the functions $\hat{f}$ are bounded and continuous; and property Q holds.

THEOREM 2. *Any partially ordered abelian group G satisfying* I, II, III, IV' *is isomorphic with an additive group of bounded continuous, real-valued functions on a completely regular space S (namely the space Σ of maximal singular subgroups of G) satisfying $P_1 - P_5$ and Q.*

Conversely, let G be an additive group of bounded functions on a set S, and let G satisfy $P_1 - P_5$. Then G satisfies I, II, III, IV'. *There is a one-to-one correspondence between S and Σ whereby the maximal singular subgroup M_{x_0} corresponding to the point x_0 of S consists of all $f \varepsilon G$ with $f(x_0) = 0$. If S is a topological space, if the functions in G are continuous, and if Q holds, then this correspondence is a homeomorphism.*

Proof. All of the first part has been shown except the validity of P_4 and P_5. P_4 is clear from the fact (KF, Lemma 2.2, p. 411) that every non-Archimedean element is contained in some maximal singular subgroup of G.

To show P_5, let Φ have the stated property, and let Ψ be the additive semigroup generated by Φ. If $f \varepsilon \Psi$ then $f = c_1 f_1 + \cdots + c_n f_n$ where each $f_i \varepsilon \Phi$ and each c_i is a positive integer. By the hypothesis on Φ, the functions f_i have a common zero M_0, i. e. $f_i \varepsilon M_0$. Then $f \varepsilon M_0$ also. Thus every element of Ψ is non-Archimedean, and the same holds for the set Ψ^* of all $f^* \varepsilon G^+$ satisfying $f^* \leqq f$ for some $f \varepsilon \Psi$. The difference group H of Ψ^* is therefore singular, and every such is contained in a maximal singular subgroup M_1 of G (KF, Lemma 2.2, p. 411). Since $\Phi \subseteq M_1$ we have $\hat{f}(M_1) = 0$ for every $f \varepsilon \Phi$, i. e. M_1 is a common zero of Φ.

Passing to the converse, let G be an additive group of bounded real functions on a set S, and let G satisfy P_1—P_5. Conditions I, II, III are obvious. Before proving IV′, we shall establish the one-to-one correspondence $x_0 \leftrightarrow M_{x_0}$ between S and Σ. M_{x_0} is a singular subgroup of G by P_3, and the obvious fact that it contains no Archimedean element. It is evidently convex. In fact it is maximal convex. For if $f \not\varepsilon M_{x_0}$, say $f(x_0) = \alpha \neq 0$, then $f - \alpha e \varepsilon M_{x_0}$. Any convex subgroup H of G containing M_{x_0} and f will therefore contain the Archimedean element $|\alpha| \cdot e$, and so will coincide with G. M_{x_0} is maximal singular. For any maximal singular subgroup of G containing M_{x_0} will be convex (KF, Lemma 2.4, p. 412), and hence will coincide with M_{x_0}. From P_2 it is clear that $x_1 \neq x_2$ implies $M_{x_1} \neq M_{x_2}$.

On the other hand, let M be any maximal singular subgroup of G, and let $M^+ = M \cap G^+$. Let $f_1, \cdots, f_n \varepsilon M^+$. Then $f_1 + \cdots + f_n = f \varepsilon M^+$. Since M contains no Archimedean elements, it follows from P_4 that $f(y_0) = 0$ for some $y_0 \varepsilon S$. From $f_1(y_0) + \cdots + f_n(y_0) = 0$ and $f_i(y_0) \geqq 0$ $(i = 1, \cdots, n)$ we conclude $f_i(y_0) = 0$. Thus any finite subset $f_1, \cdots, f_n$ of M^+ has a common zero in S. Applying P_5 we conclude the existence of $x_0 \varepsilon S$ such that $f(x_0) = 0$ for every $f \varepsilon M^+$. Since, by definition of singular subgroup, every element of M is the difference of two elements of M^+, we conclude that $M \subseteq M_{x_0}$. Then $M = M_{x_0}$ from the maximality of M and the singularity of M_{x_0}.

IV′ is now immediate since every maximal singular subgroup of G is some M_{x_0}, and the latter has been shown to be maximal convex. The last assertion of the theorem is immediate from conclusion (3) of Theorem 1.

We remark that if S is compact, then Ky Fan has shown that P_4, P_5,

and Q are consequences of P_1, P_2, P_3 and the continuity of the functions in G, G being then a characterizing subgroup of $C(S)$. P_4 follows from his Lemma 2. 2 and 5. 2. P_5 is evident from the compactness of S and the fact that the set of common zeros of a set of continuous functions is a closed set. Q is just Lemma 5. 5, together with the observation that f can be chosen to be positive. We conclude with a corollary which reduces to Ky Fan's Theorem 3 when S is compact.

COROLLARY. *Let G_1 and G_2 be additive groups of bounded continuous real-valued functions defined on topological spaces S_1 and S_2 resp., both groups satisfying P_1—P_5 and Q. If G_1 and G_2 are isomorphic then S_1 and S_2 are homeomorphic. In particular, if one is compact, so is the other.*

Proof. By Theorem 2, S_i is homemorophic with the space Σ_i of maximal singular subgroups of G_i $(i = 1, 2)$. But the topology in each Σ_i is defined by purely algebraic means, so that the one-to-one correspondence between Σ_1 and Σ_2 induced in the obvious way by an isomorphism between G_1 and G_2 is a homeomorphism. We thus have $S_1 \sim \Sigma_1 \sim \Sigma_2 \sim S_2$.

3. In all of the following examples, S denotes the discrete space of the natural numbers $1, 2, \cdots$, and S^* the space S with the limit point ∞ adjoined. S is completely regular and S^* is compact. The additive group $BC(S)$ of bounded continuous real functions on S is just that of all bounded real sequences, and $BC(S^*) = C(S^*)$ is that of all convergent real sequences $\{f(n)\}$ with $f(\infty) = \lim_{n\to\infty} f(n)$.

Example 1 is a characterizing subgroup of $C(S^*)$ not satisfying Ky Fan's Condition IV (KF, p. 423). In fact it does not possess any finite subset such that every maximal singular subgroup contains at least one of them.

Example 2 is a subgroup G of $BC(S)$ satisfying P_1—P_5 and Q. By Theorem 2 it must satisfy IV′. But from the Corollary thereto, since S is not compact, G cannot be isomorphic with a characterizing subgroup of $C(\Omega)$ for any compact Ω.

The remaining three examples show the independence of conditions P_4 and P_5 and Q under the assumption of P_1, P_2, P_3 and the continuity of the functions in G. To see the complete independence we may argue as follows.

Suppose the space S decomposes into two open-and-closed subsets S_1 and S_2. Then $BC(S)$ is the (cardinal) direct sum of $BC(S_1)$ and $BC(S_2)$. Let G_i be a subgroup of $BC(S_i)$, $i = 1, 2$, and let G be their direct sum; G is of

course a subgroup of $BC(S)$. If P is any one of the properties $P_1 — P_5$ or Q, then one readily sees that G has the property P if and only if both G_1 and G_2 have it. For example, the direct sum of Examples 3 and 4 gives an example in which P_4 and P_5 are false while Q is true. A simpler example for this case is the group of all convergent sequences (applied to S and not S^*). Likewise, for the case P_4 true, P_5 and Q both false, we may cite the group of all periodic sequences. But the author did not find any such easy examples for the two remaining cases.

Example 1. Let $a_1 = (1, 1/2, 1/3, 1/4, 1/5, \cdots)$, $a_2 = (1, 1, 1/2, 1/3, 1/4, \cdots)$, $a_3 = (1, 1, 1, 1/2, 1/3, \cdots), \cdots$ and as usual $e = (1, 1, 1, 1, 1, \cdots)$. Let G be the group of all sequences expressible as a finite linear combination

$$f = \lambda e + \rho_1 a_1 + \rho_2 a_2 + \cdots + \rho_n a_n \tag{1}$$

with real coefficients $\lambda, \rho_1, \rho_2, \cdots, \rho_n$. f is convergent with limit λ. We agree that $f(\infty) = \lim_{k\to\infty} f(k)$ for all $f \in G$ and G becomes thereby a subgroup of $C(S^*)$. We are to show that $P_1 — P_3$ hold, but IV does not. P_1 and P_2 are obvious.

To show P_3, let us first consider the point $x_0 = \infty$ of S^*. If f, as given by (1), vanishes at ∞ we must have $\lambda = 0$. But then

$$g = |\rho_1| a_1 + |\rho_2| a_2 + \cdots + |\rho_n| a_n$$

also vanishes at ∞ and has the properties $g \geqq 0$, $g \geqq f$. Now let $x_0 = k \in S$, and let $b_1 = 2e — 2a_1$, $b_k = 3e + 2a_{k-1} — 4a_k$ if $k > 1$. It is easily seen that $b_k(k) = 0$ while $b_k(j) \geqq 1$ for all $j \neq k$ (including $j = \infty$). If $f \in G$ and $f(k) = 0$, it is clear that some positive multiple of b_k will serve for g in P_3.

Hence P_3 hold, and G is a characterizing subgroup of $C(S^*)$. By Ky Fan's Lemma 5. 3, the maximal singular subgroups of G are simply the sets M_k of elements of G vanishing at $k \in S^*$. Consequently the functions $f_1, \cdots, f_n$ in Condition IV are to be sequences such that for each $k \in S^*$ there is at least one of the f_i such that $f_i(k) = 0$. The existence of such a set will be seen to be impossible when we show that no element $f \neq 0$ of G has more than a finite number of zeros in S.

Let f be given by (1), and suppose, by way of contradiction, that $f(k) = 0$ for infinitely many points k of S. Then surely $f(k_i) = 0$ for $n + 1$ points k_i satisfying $n < k_0 < k_1 < k_2 < \cdots < k_n$. Now for $k > n$,

$$f(k) = \lambda + \rho_1/k + \rho_2/(k — 1) + \cdots + \rho_n/(k — n + 1).$$

Then $f(k_i) = 0$ $(i = 0, 1, \cdots, n)$ is a system of $n + 1$ linear homogeneous

equations for $\lambda, \rho_1, \rho_2, \cdots, \rho_n$ with a determinant the value of which is readily calculated to be the ratio of

$$\prod_{i=1}^{n}(n-i)\,!\prod_{i>j}(k_i-k_j) \text{ to } \prod_{i=0}^{n} k_i(k_i-1)\cdots(k_i-n+1),$$

and therefore distinct from 0. Hence $\lambda=\rho_1=\rho_2=\cdots=\rho_n=0$, that is, $f=0$.

It might be objected that Condition IV could hold vacuously, in the sense that every maximal convex subgroup of G is also maximal singular. We show that this can not happen for any characterizing subgroup G of $C(\Omega)$ with a compact Ω consisting of more than one point. Let x_1, x_2 be distinct points of Ω, and define for each $f \varepsilon G$, $\tau(f)=\frac{1}{2}f(x_1)+\frac{1}{2}f(x_2)$. τ is a homomorphism of G into the simply ordered additive group of real numbers, and, by KF, Lemma 3.4, the set M of all $f \varepsilon G$ satisfying $\tau(f)=0$ is a maximal convex subgroup of G. Were M singular, every element f of M would be the difference $f=g-h$ of two positive elements g, h of M. But $g \geqq 0$ and $\tau(g)=0$ imply $g(x_1)=g(x_2)=0$, and similarly for h. Hence $f(x_1)=f(x_2)=0$. Thus M would be contained in the two distinct maximal convex subgroups M_{x_1} and M_{x_2} of G, contrary to its maximality.

Example 2. Let H be the group of all essentially finite sequences $(\alpha_1, \alpha_2, \cdots, \alpha_n, 0, 0, \cdots)$ with $\alpha_1+\alpha_2+\cdots+\alpha_n=0$. Evidently H contains no properly positive sequence. Let G consists of all sequences $\lambda e+h$ with λ any real number and $h \varepsilon H$. We proceed to show P_1—P_5 and Q, regarding G as a subgroup of $BC(S)$. P_1 and P_2 are evident.

Both P_3 and Q are evident when we exhibit, for each $k \varepsilon S$, an element b_k of G such that $b_k(k)=0$, $b_k(j) \geqq 1$ for all $j \neq k$. But such an element is $e+h_k$, where $h_k(j)=-1$ if $j=k$, $h_k(j)=1$ if $j=k+1$, and $h_k(j)=0$ for all other j.

To show P_4, suppose that $f=\lambda e+(\alpha_1, \alpha_2, \cdots \alpha_n, 0, 0, \cdots)$, $\sum_i \alpha_i=0$, is a non-Archimedean element of G. Were $f(k) \geqq 0$ for every $k \varepsilon S$ we would have $\lambda>0$ and $\lambda+\alpha_i>0$ for every $i=1, 2, \cdots, n$. This would contradict the condition $\inf_{k \varepsilon S} f(k)=0$ that f be non-Archimedean (KF, Lemma 5.1, p. 419).

To show P_5, let Φ be a set of positive functions of G with the stated property. Now a positive element $\lambda e+h$ of G must have $\lambda>0$, and consequently can have only a finite number of zeros in S. Let $f_0 \varepsilon \Phi$, and let $k_1, \cdots, k_r$ be its zeros in S. If the conclusion to P_5 were false there would exist in Φ for each $i=1, \cdots, r$ an element f_i of G for which $f_i(k_i) \neq 0$.

But then the finite set $f_0, f_1, \cdots, f_r$ would have no common zero, contrary to hypothesis.

It is instructive to note that G can be regarded as a subgroup of $C(S^*)$, but fails to be a characterizing subgroup thereof since $M_\infty - H$ is not singular.

Example 3 (independence of P_4). Consider the group of Example 1 as a subgroup of $BC(S)$. Properties $P_1 - P_3$ are unaffected thereby. P_4 is false since a_1 is non-Archimedean but vanishes nowhere in S. P_5 follows from the fact that each element $\neq 0$ of G has only a finite number of zeros in S. Q follows from the existence of the sequences b_k.

Example 4 (independence of P_5). Let G be the group of all eventually constant sequences, regarded as a subgroup of $BC(S)$. $P_1 - P_4$ and Q are all evident. To show the falsity of P_5 one may take for Φ the following set of sequencies: $f_1 = (1, 0, 0, 0, \cdots)$, $f_2 = (1, 1, 0. 0, \cdots)$, $f_3 = (1, 1, 1, \cdots), \cdots$.

Example 5 (independence of Q). Let T denote the set S^* endowed with discrete topology. Let $G = C(S^*)$, but regarded as a subgroup of $BC(T)$. The properties $P_1 - P_5$ are unaffected thereby, but Q is false since there exists no $f \varepsilon G$ with $f(\infty) = 0$ and $\inf_{k<\infty} f(k) > 0$.

THE JOHNS HOPKINS UNIVERSITY.

A NOTE ON LIE ALGEBRAS OF CHARACTERISTIC p.*

By N. Jacobson.

Let $\mathfrak{L}$ be a finite dimensional Lie algebra over a field of characteristic p, let $\mathfrak{A}$ be the universal associative algebra of $\mathfrak{L}$ ([1] and [4]) and let $\mathfrak{C}$ be the center of $\mathfrak{A}$. In this note we prove that if a is a linear element of $\mathfrak{A}$ then there exists a non-zero polynomial ϕ such that $\phi(a) \in \mathfrak{C}$. We use this result to obtain the following: (1) a simple direct proof of Iwasawa's theorem ([2], p. 420) that every finite dimensional Lie algebra of characteristic p has a faithful finite dimensional representation, (2) a proof of a conjecture of Chevalley that every finite dimensional Lie algebra of characteristic p has a representation which is not completely reducible, (3) a proof that $\mathfrak{A}$ can be imbedded in a division algebra.

1. A polynomial of the form $\alpha_0\lambda^{p^m} + \alpha_1\lambda^{p^{m-1}} + \cdots + \alpha_m\lambda$ is called a *p-polynomial.*

Lemma. *Any polynomial $\neq 0$ is a factor of a suitable p-polynomial $\neq 0$.*

Proof. Let $\mu(\lambda) \neq 0$. For each $i = 0, 1, 2, \cdots$ write $\lambda^{p^i} = \mu(\lambda)q_i(\lambda) + r_i(\lambda)$ where $\deg r_i(\lambda) < \deg \mu(\lambda)$. Only a finite number of the $r_i(\lambda)$ are linearly independent. Hence we can find β_i not all 0 such that $\Sigma\beta_i r_i(\lambda) = 0$. Then $\Sigma\beta_i\lambda^{p^i} = \mu(\lambda)(\Sigma\beta_i q_i(\lambda))$.

Proposition 1. *Let a be a linear element of the universal associative algebra $\mathfrak{A}$ of a finite dimensional Lie algebra $\mathfrak{L}$ of characteristic p. Then there exists a non-zero polynomial $\phi(\lambda)$ such that $\phi(a)$ is in the center $\mathfrak{C}$ of $\mathfrak{A}$.*

Proof. Let A denote the adjoint mapping $x \to [xa]$ in $\mathfrak{L}$. Since $\mathfrak{L}$ is finite dimensional there exists a non-zero polynomial $\mu(\lambda)$ such that $\mu(A) = 0$. Let $\phi(\lambda) = \Sigma\beta_i\lambda^{p^i}$ be a p-polynomial divisible by $\mu(\lambda)$. Then also $\phi(A) = 0$. For any $x \in \mathfrak{L}$ (regarded as the linear part of $\mathfrak{A}$)[1]

$$[x, \phi(a)] = \Sigma\beta_i[x, a^{p^i}] = \Sigma\beta_i[\cdots[x, \overbrace{a]\cdots a}^{p^i}] = x\Sigma\beta_i A^{p^i} = 0.$$

Hence $\phi(a) \in \mathfrak{C}$.

* Received April 28, 1951.

[1] We have used here the identity $[xa^p] = [\cdots[x\overbrace{a]\cdots a}^{p}]$. Cf. [3], p. 16.

Choose a basis $(e_1, e_2, \cdots, e_n)$ for $\mathfrak{L}$ and for each e_i let $\phi_i(\lambda)$ be a non-zero polynomial such that $y_i = \phi_i(e_i) \varepsilon \mathfrak{C}$. Let $d_i = \deg \phi_i(\lambda)$. Let $\mathfrak{B}$ be the ideal in $\mathfrak{A}$ generated by the elements y_i, $i = 1, 2, \cdots, n$.

PROPOSITION 2. *The cosets of the elements* $e_1^{\lambda_1} e_2^{\lambda_2} \cdots e_n^{\lambda_n}$, $0 \leq \lambda_i < d_i$, $(\lambda_1, \lambda_2, \cdots, \lambda_n) \neq 0$ *form a basis for* $\mathfrak{A}/\mathfrak{B}$.

Proof. It is known that the elements $e_1^{k_1} e_2^{k_2} \cdots e_n^{k_n}$, $k_i = 0, 1, 2, \cdots$, $(k_1, \cdots, k_n) \neq 0$ form a basis for $\mathfrak{A}$. Any polynomial in e_i can be expressed as a linear combination of elements $y_i^{m_i} e_i^{\lambda_i}$, $0 \leq \lambda_i < d_i$. Hence any element of $\mathfrak{A}$ is a linear combination of terms of the form $y_1^{m_1} y_2^{m_2} \cdots y_n^{m_n} e_1^{\lambda_1} e_2^{\lambda_2} \cdots e_n^{\lambda_n}$, $0 \leq \lambda_i < d_i$. It follows that every element is congruent $\bmod \mathfrak{B}$ to a linear combination of the elements $e_1^{\lambda_1} \cdots e_n^{\lambda_n}$. It remains to show that these elements are linearly independent $\bmod \mathfrak{B}$. Any element $b \varepsilon \mathfrak{B}$ is a linear combination of the terms $y_1^{m_1} y_2^{m_2} \cdots y_n^{m_n} e_1^{\lambda_1} e_2^{\lambda_2} \cdots e_n^{\lambda_n}$ where $(m_1, m_2, \cdots, m_n) \neq 0$. If we write

$$y_1^{m_1} \cdots y_n^{m_n} e_1^{\lambda_1} \cdots e_n^{\lambda_n} = y_1^{m_1} e_1^{\lambda_1} y_2^{m_2} e_2^{\lambda_2} \cdots$$

and express y_i in terms of e_i we obtain

$$y_1^{m_1} \cdots y_n^{m_n} e_1^{\lambda_1} \cdots e_n^{\lambda_n} = e_1^{\lambda_1 + m_1 d_1} e_2^{\lambda_2 + m_2 d_2} \cdots e_n^{\lambda_n + m_n d_n} + *$$

where the $*$ is of degree $< \Sigma(\lambda_i + m_i d_i)$. It follows that if $b \neq 0$, then in the expression for b in terms of the basis $e_1^{k_1} e_2^{k_2} \cdots e_n^{k_n}$ at least one of the e_i has multiplicity $\geqq d_i$. Hence no non-zero linear combination of the terms $e_1^{\lambda_1} \cdots e_n^{\lambda_n}$, $\lambda_i < d_i$ belongs to $\mathfrak{B}$.

2. As a first application of Propositions 1 and 2 we prove

THEOREM 1. (*Iwasawa*) *Any finite dimensional Lie algebra of characteristic* $p \neq 0$ *has a faithful finite dimensional representation* ([2], p. 420).

Proof. Choose $y_i = \phi_i(e_i)$ of degree $d_i > 1$ in Proposition 2. Then the cosets $\bar{e}_i = e_i + \mathfrak{B}$ are linearly independent. Hence the natural imbedding of $\mathfrak{L}$ in $\mathfrak{A}/\mathfrak{B}$ is 1-1. On the other hand $\mathfrak{A}/\mathfrak{B}$ has the finite basis $e_1^{\lambda_1} e_2^{\lambda_2} \cdots e_n^{\lambda_n} + \mathfrak{B}$, $0 \leq \lambda_i < d_i$. Hence $\mathfrak{A}/\mathfrak{B}$ and, a fortiori, $\mathfrak{L}$ has a faithful finite dimensional representation.

THEOREM 2. (*Conjecture of Chevalley*) *Any finite dimensional Lie algebra of characteristic* p *has a finite dimensional representation which is not completely reducible.*

Proof. We now take $y_i = \phi_i(e_i)$ where $\phi_i(\lambda) = [\psi_i(\lambda)]^{n_i}$, $\deg \psi_i(\lambda) > 1$,

$\psi_i(e_i) \in \mathfrak{C}$ for all i, $n_i > 1$ for some i. Then $\bar{z}_i = \psi_i(e_i) + \mathfrak{B}$ is a non-zero element in the center of $\mathfrak{A}/\mathfrak{B}$. On the other hand, $\bar{z}_i^{n_i} = 0$ so that the center of $\mathfrak{A}/\mathfrak{B}$ has nilpotent elements. Hence the finite dimensional algebra $\mathfrak{A}/\mathfrak{B}$ is not semi-simple and therefore it has a finite dimensional representation which is not completely reducible. The same holds for $\mathfrak{L}$.

THEOREM 3. *The universal associative algebra of any finite dimensional Lie algebra of characteristic p can be imbedded in a division algebra.*

Proof. We can construct in the usual way the ring Δ of quotients ac^{-1}, $a \in \mathfrak{A}$, $c \neq 0$ in $\mathfrak{C}$. Δ is an integral domain and contains the quotient field Γ of $\mathfrak{C}$ in its center. If $(e_1, e_2, \cdots, e_n)$ is a basis for $\mathfrak{L}$ and $\phi_i(e_i) \in \mathfrak{C}$ where $\deg \phi_i = d_i > 0$ then we have seen that every element of $\mathfrak{A}$ is a linear combination with coefficients in $\mathfrak{C}$ (polynomials in $y_i = \phi_i(e_i)$) of the elements $e_1^{\lambda_1} e_2^{\lambda_2} \cdots e_n^{\lambda_n}$, $0 \leq \lambda_i < d_i$. It follows that Δ is finite dimensional over Γ. Hence Δ is a division algebra.

YALE UNIVERSITY.

REFERENCES.

[1] G. Birkhoff, "Representability of Lie algebras . . .," *Annals of Mathematics*, vol. 38 (1937), pp. 326-332.

[2] K. Iwasawa, "On the representation of Lie algebras," *Japanese Journal of Mathematics*, vol. 19 (1948), pp. 405-426.

[3] N. Jacobson, "Restricted Lie algebras of characteristic p," *Transactions of the American Mathematical Society*, vol. 50 (1941), pp. 15-25.

[4] E. Witt, "Treue Darstellung Liescher Ringe," *Journal für die reine und angewandedte Mathematik*, vol. 177 (1937), pp. 152-160.

ON THE LOGARITHMS OF BOUNDED MATRICES.*

By AUREL WINTNER.

Let capital letters denote (infinite) bounded matrices of complex numbers. Such a matrix, say A, is called non-singular if it has a unique (right- *and* left-sided) bounded reciprocal matrix. If the spectrum, sp A, of a bounded A is defined to be the set of those complex numbers λ for which the matrix $\lambda I - A$ fails to be non-singular, then, as proved by me some time ago,[1]

(1) $$\text{sp}\, A \text{ is not vacuous}$$

(*loc. cit.*, p. 243) and

(2) $$\text{sp}\, A \text{ is closed and bounded}$$

(*loc. cit.*, p. 242), finally

(3) $$\text{sp}\, A \text{ is in } \{A\}$$

(*loc. cit.*, p. 245). Here $\{A\}$ denotes the closure of the values attained by the bounded bilinear form, $A(x, y)$, of A under the following pair of restrictions: $y = \bar{x}$ and $|x| = 1$, where $\bar{x} = (\bar{x}_1, \bar{x}_2, \cdots)$ and $|x| = (|x_1|^2 + |x_2|^2 + \cdots)^{\frac{1}{2}}$, if $x = (x_1, x_2, \cdots)$. Corresponding to the Toeplitz-Hausdorff theorem on finite matrices,

(4) $$\{A\} \text{ is convex.}$$

As pointed out *loc. cit.* (pp. 246-249 and p. 244, respectively), neither (1) nor (3) can be concluded, by a limit process, from the corresponding facts on finite matrices, whereas (4) readily follows by such a process.

If A is any set of points in the complex plane, let A^* denote its complement. Then (2) and (1) show that $(\text{sp}\, A)^*$ is an open set containing a domain $a < |\lambda| < \infty$ (if a is large enough) but not containing every point of the plane. Let $\mathrm{A}^0, \mathrm{A}^1, \cdots$ be the (finite or infinite) sequence representing the components of the open set $(\text{sp}\, A)^*$, so that

(5) $$\text{sp}\, A = (\sum_n \mathrm{A}^n)^*,$$

* Received June 28, 1951.

[1] A. Wintner, "Zur Theorie der beschränkten Bilinearformen," *Mathematische Zeitschrift*, vol. 30 (1929), pp. 228-282.

where every A^n is a connected open set which, if $m \neq n$, contains no point of A^m. Let the notation be so chosen that A^0 is the (unique) unbounded component of $(\mathrm{sp}\, A)^*$; so that

$$(6) \qquad \sum_{n \neq 0} \mathrm{A}^n \text{ is bounded}$$

(possibly vacuous) and, according to (3),

$$(7) \qquad \{A\}^* \text{ is in } \mathrm{A}^0.$$

The following considerations will contribute to the question of the existence of a bounded matrix function $f(A)$ of a given A. While the method will be presented in the important case $f(z) = \log z$ only, its applicability to more general functions $f(z)$ will be obvious. Incidentally, since it will consist of a process of analytic continuation, it will also apply to certain linear spaces more general than Hilbert's space.

(i) Corresponding to every component A^n of the complement $\sum \mathrm{A}^n$ of the spectrum of a bounded matrix A, there exists a bounded, non-singular matrix C_n which is commutable with A and has the following property: There belongs to every point λ of A^n a bounded matrix A_λ which is commutable with A and satisfies the relation

$$(8) \qquad \lambda I - A = C_n \exp A_\lambda.$$

In addition, A_λ can be chosen so as to be continuous (and, for that matter, analytic and regular) on the λ-domain A^n.

The latter notions refer to that metrization of the space of all bounded matrices A in which the norm $|A|$ is defined as the least upper bound of the vector length $|Ax|$ when x varies over all vectors of unit length.

Since $e^B e^C = e^{B+C}$ when B, C are bounded and commutable, it is clear that (i) contains the following corollary:

(ii) With reference to every fixed component A^n of $(\mathrm{sp}\, A)^*$, the matrix $\lambda I - A$ has a bounded logarithm either for every or for no λ contained in A^n.

It is understood that a bounded matrix C is called a logarithm of a bounded matrix if the latter is e^C. Since $e^C e^{-C} = I = e^{-C} e^C$, only a non-singular matrix can have a bounded logarithm. If not every non-singular matrix can be represented as the square of a bounded matrix, then not every non-singular matrix has a bounded logarithm.

7

If $|Z| < 1$, then $I + Z$ has a bounded logarithm. In fact, one such logarithm is supplied by the series

$$\log(I + Z) = \sum_{m=1}^{\infty} m^{-1}(-Z)^{m+1} \text{ if } |Z| < 1. \tag{9}$$

This implies that if A is bounded and if the absolute value of λ is large enough, then $\lambda I - A$ has a bounded logarithm (in fact, (9) is applicable to $Z = -\lambda^{-1}A$ if $|\lambda| > |A|$). Since the matrix (9) is commutable with Z, it follows that (ii) contains the following corollary:

(iii) If λ is in $\mathbf{A}^0$, i. e., in the unbounded component of the complement of $\operatorname{sp} A$ (e. g., in the complement of $\{A\}$), then $\lambda I - A$ has a bounded logarithm.

The parenthetical assertion of (iii) follows, by (3), from the main assertion of (iii).

If A is completely continuous, then $\operatorname{sp} A$ consists of the point $\lambda = 0$ and of a finite (possibly vacuous) or infinite sequence of points which can cluster only at $\lambda = 0$ (Hilbert). Since this implies that $(\operatorname{sp} A)^*$ must be a connected set, and therefore identical with $\mathbf{A}^0$, the first part of the following assertion is a consequence of (iii):

(iv) If A is completely continuous, then $\lambda I - A$ has a bounded logarithm for every λ not contained in $\operatorname{sp} A$. In addition, this logarithm can always be chosen so as to be completely continuous.

The second part of (iv) will be clear from the proof of (i); cf. (II) below.

The proof of (i) centers about the following fact:

(I) If A and B are two bounded, commutable matrices the first of which is non-singular while the second is such as to satisfy

$$|B| < |A^{-1}|^{-1}, \tag{10}$$

then there exists a bounded matrix C which is commutable with A and with B and which satisfies the relation

$$A + B = Ae^{C}. \tag{11}$$

In fact, since $A + B$ is the product of A and $I + A^{-1}B = I + BA^{-1}$ (in either order), a C satisfying all the requirements of (I) results by chosing

$$Z = A^{-1}B \tag{12}$$

in (9). For, since (10) and (12) imply that $|Z| \leqq |A^{-1}|\,|B| < 1$, the proviso $|Z| < 1$ of (9) is satisfied.

Since the product of a bounded and of a completely continuous matrix is completely continuous, it is clear from this proof of (I), and from (12), that

(II) C in (I) can be chosen to be completely continuous if B is completely continuous.

It is easy to conclude from (I) the following lemma:

(III) Let A be a non-singular matrix corresponding to which there exists a family of non-singular matrices $A(t)$, $0 \leqq t \leqq 1$, satisfying the pair of conditions

$$A(0) = I, \qquad A(1) = A \tag{13}$$

and the commutability condition

$$A(u)A(v) = A(v)A(u). \tag{14}$$

Suppose further that

$$\underset{0 \leqq t \leqq 1}{\text{g. l. b.}}\, |A^{-1}(t)|^{-1} > 0 \tag{15}$$

and that $A(t)$ depends on t continuously. Then there exists a bounded matrix C satisfying $A = e^C$.

By the continuity of $A(t)$ is meant that, if ϵ is positive, and if $\delta = \delta_\epsilon > 0$ is small enough, then $|A(u) - A(v)| < \epsilon$ whenever $|u - v| < \delta$. Hence, if a is any positive constant, and if $j = j_a$ is large enough, then there exist $j + 1$ values $0 = t_0 < t_1 < \cdots < t_j = 1$ satisfying

$$|A(t_k) - A(t_{k-1})| < a \text{ for } k = 1, \cdots, j. \tag{16}$$

Choose a to be the greatest lower bound (15). Then it is seen from (14) and (13), where $A(0) = \exp(0)$, that (III) follows by j consecutive applications of (I), namely, by identifying the matrices B and A in (I) with the matrices $A(t_k) - A(t_{k-1})$ and $A(t_{k-1})$, respectively.

It is readily seen that (i), the first of the assertions (i)-(iv) made above, is a corollary of (III). In fact, $\lambda I - A$ and $\mu I - A$ are always commutable, as are the reciprocals of $\lambda_1 I - A$ and $\lambda_2 I - A$ if λ_1 and λ_2 are not in sp A. On the other hand, since every component $\mathbf{A}^n$, to which (i) refers, is an open connected set in the λ-plane, the existence of an appropriate continuous family $A(t)$, $0 \leqq t \leqq 1$, having the properties specified in (III), is obvious.

Remark. Suppose that A is a finite matrix. Then it has a logarithm, i. e., $A = e^C$ holds for some C, if (and only if) $\det A \neq 0$. In other words, $\lambda I - A$ has a logarithm whenever $\phi(\lambda) \neq 0$, where $\phi(\lambda) = \det(\lambda I - A)$. The standard proof of this theorem depends either on the existence of the Weierstrass-Jordan normal form of A, supplied by the theory of elementary divisors, or [2] (somewhat more directly, though in the main equivalently) on the Cayley-Hamilton theorem, $\phi(A) = 0$. The above considerations contain a proof which does not involve any such machinery. In fact, since sp A now consists of a finite number of points, the complement of sp A is A^0, hence the proof of (iii) or (iv) is applicable.

THE JOHNS HOPKINS UNIVERSITY.

[2] L. Schlesinger, "Über den Logarithmus einer Matrix," *Journal für die reine und angewandte Mathematik*, vol. 161 (1929), pp. 199-200. Cf. G. Pólya, "Der Picardsche Satz für Matrizen," *Jahresbericht der Deutschen Mathematiker Vereiningung*, vol. 40 (1931), second part, p. 80.

ON PARALLEL SURFACES.*

By Aurel Wintner.

1. Let $S: X = X(u, v)$, where $X = (x, y, z)$, be a surface of class C^2. Then, if S is sufficiently small, it can be assumed to be given in the form $z = z(x, y)$, where z has continuous partial derivatives of first and second order, $p = z_x, \cdots, t = z_{yy}$, and the total and mean curvatures, $K = K(x, y)$ and $H = H(x, y)$, of S are given by

$$(1) \qquad Kd^4 = rt - s^2, \text{ where } d = (1 + p^2 + q^2)^{\frac{1}{2}},$$

and

$$(2) \qquad Hd^3 = \tfrac{1}{2}(1 + p^2)t - pqs + \tfrac{1}{2}(1 + q^2)r.$$

The sign of the first, but not of the second, of the continuous functions K, H has a geometrical meaning (H contains an odd power of the square root of the two-valued function $\pm d$).

Suppose that S is such as to make $K(x, y)$ a positive constant, say 1. Then (2) reduces to

$$(3) \qquad z_{xx}z_{yy} - z_{xy}^2 = (1 + z_x^2 + z_y^2)^2,$$

a partial differential equation of second order which, according to S. Bernstein's general theorem on analytic differential equations of elliptic type, has the property of possessing no C^3-solution $z = z(x, y)$ which is not analytic as well. Accordingly, every surface of class C^3 having a constant positive curvature K is an analytic (for a more inclusive geometrical application of Bernstein's general theorem, cf. [1] and the comments in [6], Section 3).

This result contains, however, an unsatisfactory element which, along the lines of the preceding argument, would disappear only if it were known that Bernstein's general theorem (on arbitrary analytic differential equations of elliptic type in a plane) remains true for all solutions of class C^2. In fact those statements in differential geometry which require more than the mere existence of continuous derivatives of second order can hardly be justified for geometrical reasons. The restriction to the class C^3 (or, for that matter, $C^2(\lambda)$, where λ is a Hölder index; cf. [6], Section 5) often proves to be of an

* Received February 27, 1951.

artificial or preliminary nature, that is to say, such as to be necessitated by the lack of appropriate analytical tools only. Suffice it to observe that every surface of class C^2 has continuous first *and* second fundamental forms, and therefore, in particular, continuous curvatures. It is therefore of more than just formal interest that the above-mentioned consequence of Bernstein's general theorem can be freed of *all* of its non-geometrical restrictions, as follows:

(i) *If the total curvature of a surface of class C^2 is a positive constant, then the surface must be analytic.*

The point in this theorem is that, in contrast to the lemma of Section 3 (below) on which part of the proof of the theorem will depend, no data involving the second derivatives are subject to a Hölder restriction.

Needless to say, the analogue of the theorem is false if $K = \text{const.} \leqq 0$. This is obvious in the case const. $= 0$, since a torse can clearly be of class C^∞ without being analytic. Similarly, if const. $= -1 < 0$, then, since (1) becomes an analytic differential equation of hyperbolic type, it will surely possess non-analytic solutions $z = z(x, y)$ of class C^∞.

2. If $z(x, y)$ is a function possessing continuous second derivatives which are not subject to any restriction, then nothing but mere continuity can be claimed for the functions (1), (2). Similarly, if the second derivatives of $z(x, y)$ satisfy a Hölder condition (of some index λ), then the same is true (with the same λ) of both functions (1), (2). This makes clear the nature of the content of the following lemma:

(ii) *If a function $z = z(x, y)$ is of class C^2, and if its derivatives $p = z_x(x, y), \cdots, t = z_{yy}(x, y)$, besides being continuous, have the property that their combination $H = H(x, y)$, defined by (2), possesses continuous first derivatives H_x, H_y and that the latter satisfy, for some pair of positive constants γ, λ, a Hölder condition*

$$(4) \qquad | H_x' - H_x'' | \leqq \gamma R^\lambda, \qquad | H_y' - H_y'' | \leqq \gamma R^\lambda,$$

where $H_x' = H_x(x', y'), \cdots$ and $R^2 = (x' - x'')^2 + (y' - y'')^2$, then $z = z(x, y)$ must be of class C^3.

In particular, the function $K = K(x, y)$, being defined by (1), must then be of class C^1.

(ii bis) *The assertion of* (ii) *becomes false if $H(x, y)$ in the assumption* (4) *is replaced by $K(x, y)$.*

In order to see truth of this negation, (ii bis), let $z(x, y) = f(x + y)$, where $f(t)$ is a function of a single variable and has a continuous second derivative. Then $S: z = z(x, y)$ is of class C^2 and the function $K = K(x, y)$, being defined by (1), vanishes identically. Hence, (ii bis) follows from the circumstance that a function $f(t)$ possessing a continuous second derivative need not have a third derivative. (Nothing like this can, however, occur if the partial differential equation (1) for $z(x, y)$ is of elliptic type, that is, if $K(x, y) > 0$.)

3. The proof of (ii) proceeds as follows:

Let E_3 denote the three-dimensional Euclidean (z, p, q)-space, F_2 a (connected, open) domain in the (x, y)-plane, and L a function defined on the product space of E_3 and F_2 so as to satisfy the following conditions: If (x, y) is any fixed point of F_2, then L is fairly smooth (say, analytic) on E_3 and has a Hessian $L_{pp}L_{qq} - L_{pq}{}^2$ which is positive throughout; while, as a function of all five variables together, L is of class $C^1(\lambda)$, for some Hölder index $\lambda < 1$ (by virtue of the first of these conditions, the $C^1(\lambda)$-condition means that (4) is satisfied by $D = L$ and by a γ which can be chosen independent of (z, p, q) on every bounded portion of E_3). Under these assumptions, if $[L]$ denotes the Lagrangian derivative of L, every function $z = z(x, y)$ which is of class C^2 and satisfies the differential equation $[L] = 0$, on F_2, must be of class C^3. This follows from the proof, even though not from the wording, of a theorem of Lichtenstein [4]. In fact, although he assumes the analyticity of L in all five variables together, a glance at the proof shows what is actually needed is contained in the above conditions.

Let $H(x, y)$ denote a given function of class $C^1(\lambda)$ on an (x, y)-domain F_2. Then the conditions just required of L are satisfied by either of the functions

$$L(z, p, q; x, y) = d - H(x, y)z, \text{ where } d = \pm(1 + p^2 + q^2)^{\frac{1}{2}}.$$

But the Lagrangian equation $[L] = 0$ for a $z = z(x, y)$ then becomes precisely the condition (2), where $p = z_x, \cdots, t = z_{yy}$. Hence every function $z = z(x, y)$ of class C^2 which satisfies the relation (2), belonging to any given $H = H(x, y)$ of class $C^1(\lambda)$, must be a function of class C^3. This proves (ii).

4. For a given $n \geqq 1$, and for a sufficiently small piece, S, of a surface of class C^n, let

$$(5) \qquad S: X = X(u^1, u^2),$$

where $X = (x, y, z)$, be a C^n-parametrization of S. By this is meant that, in the small (u^1, u^2)-domain under consideration, the (vector) function (5), besides being of class C^n, is such that the vector product $[X_1, X_2]$ of the partial derivatives $X_i = \partial X/\partial u^i$ does not vanish. Thus there exists on S a normal vector of unit length,

$$(6) \qquad N = [X_1, X_2]/|[X_1, X_2]|,$$

which is a function $N(u^1, u^2)$ of class C^{n-1}. Hence, if c is a given number distinct from 0, then the (u^1, u^2)-parametrization of the X-set $S^* = S^*(c)$ which, in the sense of Steiner [5], is "the parallel surface of S at the distance c" (on one or on the other side of S, according as $c > 0$ or $c < 0$), that is, the parametrization

$$(7) \qquad S^*\colon X = X^*(u^1, u^2) = X(u^1, u^2) + cN(u^1, u^2)$$

of the "parallel set" $S^* = S^*(c)$, is based on a function, $X^*(u^1, u^2)$, which is of class C^{n-1}. In fact, if the function (5) is of class C^n, then, in view of the differentiations contained in (6), the given function (7) will not in general be of class C^n. The following fact is therefore quite unexpected:

(iii) *If a surface S is of class C^2 and is sufficiently small with reference to a given real number c, then the corresponding parallel surface, $S^* = S^*(c)$, of S is again of class C^2, provided that* (with reference to the numerical values, K_0 and H_0, of the curvatures at that point of S the immediate vicinity of which is considered) *two numerical values of c are excepted*; *cf.* (19) *below.*

This does not mean, of course, that, barring the excepted c-values, (7) is a local C^2-parametrization whenever (5) is, but that (7) must then possess *some* local C^2-parametrizations,

$$(7\text{ bis}) \qquad S^*\colon X = X^*(v^1; v^2),$$

in terms of *certain* parameters v^1, v^2 (which, in general, must be distinct from the given u^1, u^2). On the other hand, *both* $z(x, y)$ and $z^*(x, y)$ in

$$(8) \qquad S\colon z = z(x, y) \quad \text{and} \quad S^*\colon z = z^*(x, y)$$

must be functions of class C^2 in some circle $x^2 + y^2 < b = b(a; c)$ if the local C^2-parametrization (5) of S is replaced by the Cartesian representation $z = z(x, y)$ in a circle $x^2 + y^2 < a$ the plane of which is the plane tangent to S at $(x, y) = (0, 0)$. In fact, the normal of S at $(0, 0)$ then becomes the z-axis, and the same is true of the normal of S^* at $(0, 0)$. It follows therefore from (iii), and from a standard application of the C^2-form of the

classical theorem on implicit functions, that both functions (8) must be of class C^2 in a vicinity of $(x, y) = (0, 0)$.

5. The actual content of (iii) is revealed by the second part $(n = 1)$ of the following assertion:

(iii*) *The C^2 in* (iii) *can be replaced by C^n if $n \geqq 2$ but not if $n = 1$.*

While the first assertion of (iii*) will be clear from the proof (iii) below, failure for $n = 1$ can be concluded from the following example: For small $|x|$, let $f(x)$ be a function possessing a continuous derivative, $f'(x)$, and let $f(0) = 0$ and $f'(0) = 0$. Then, if $z(x, y)$ is defined to be $f(x)$ for all y, the surface S: $z = z(x, y)$ is of class C^1 and, since $f'(0) = 0$, the plane $z = 0$ is tangent to S along the y-axis. Hence, if a parallel surface $S^* = S^*(c)$ of $S: z = f(x)$ is of class C^1, then it must be representable, for small $|x|$, in the form $z^* = f^*(x)$, where $f^*(x)$ has a continuous first derivative which vanishes at $x = 0$. But it turns out that such is not the case if $f(x)$ is suitably chosen.

First, since S and S^* are represented by $z = f(x)$ and $z^* = f^*(x)$, it is seen from (6) and (7), where $X = (x, y, z) = (u^1, u^2, f)$ and $X^* = (u^1, u^2, f^*)$, that the function $z^* = f^*(x)$ is the result of the elimination of x between the two equations

$$x = u - f'(u)(1 + f'^2(u))^{-\frac{1}{2}}, \qquad z^* = f(u) + (1 + f'^2(u))^{-\frac{1}{2}},$$

if, without loss of generality, $c = 1$ in $S^* = S^*(c)$. Here $f'(u)$ is any continuous function vanishing at $u = 0$, and $f(u)$ is its primitive vanishing at $u = 0$. Accordingly, if $f' = g$, then for small $|u|$,

$$x = u - g(u)(1 + g^2(u))^{-\frac{1}{2}}, \tag{9}$$

where $g(u)$ is any continuous function satisfying $g(0) = 0$, and

$$z^* = (1 + g^2(u))^{-\frac{1}{2}} + \int_0^u g(v)\,dv. \tag{10}$$

The assertion is that, if $g(u)$ is suitably chosen, and if $z^* = f^*(x)$ denotes, for small $|x|$, the (continuous) function defined by the parametric equations (9)-(10), then $f^*(x)$ can fail to have a continuous derivative which vanishes at $x = 0$.

The square root, $(\quad)^{-\frac{1}{2}}$, in (9)-(10) must be chosen positive if, corresponding to the choice $c = 1$, the value of the ordinate $z^* = f^*(x)$ at $x = 0$

is normalized to be $+1$. Then, since $(1+g^2)^{-\frac{1}{2}} = 1 - \frac{1}{2}g^2 + o(g^2)$ as $u \to \pm 0$, it follows that, in view of (10), the function $z^* = f^*(x)$ must satisfy the relation

$$(11) \qquad f^*(x) - f^*(0) = -\tfrac{1}{2}g^2(u) + o(g^2(u)) + \int_0^u g(v)\,dv$$

by virtue of the relation (9), and that the latter is of the form

$$(12) \qquad x = u - g(u) + o(|\,g(u)\,|^3).$$

Here $g(u)$ can be chosen to be any continuous function which vanishes at $u = 0$. But it is not hard to construct a (sufficiently "wobbly") continuous function $g(u) = o(1)$ which has the property that the ratio of (11) to (12) fails to tend to a limit as $u \to 0$ (even if u-values at which the denominator, (12), vanishes, are disregarded; these u-values can be confined to a sequence which clusters only at $u = 0$). For such a $g(u)$, the ratio of $f^*(x) - f^*(0)$ to x fails to tend to a limit as $x \to 0$. Since this means that $f^*(x)$ is non-differentiable at $x = 0$, the truth of the negative assertion of (iii*) follows.

6. In order to prove (iii), suppose that S is of class C^2 (and that (5) is a C^2-parametrization of S). Then S, besides having a first fundamental form $g_{\alpha\beta}du^\alpha du^\beta$ in which the coefficients

$$(13) \qquad g_{ik} = X_i \cdot X_k = g_{ki} \qquad (\det g_{ik} > 0)$$

are functions of class C^1 in (u^1, u^2), possesses a second fundamental form $h_{\alpha\beta}du^\alpha du^\beta$ in which the coefficients

$$(14) \qquad h_{ik} = -X_i \cdot N_k = h_{ki} \qquad (h_{ik} = N \cdot X_{ik})$$

are continuous (possibly non-differentiable) functions of (u^1, u^2), the vector function (6) being of class C^1. The first derivatives of (6) are connected with those of (5) by Weingarten's identities

$$(15) \qquad N_i = -h_{i\alpha}g^{\alpha\beta}X_\beta,$$

where, in matrix notations, $(g^{ik}) = (g_{ik})^{-1}$. In terms of (13) and (14), the representations of the total and mean curvatures, K and H, are

$$(16) \qquad K = (\det h_{ik})/(\det g_{ik}), \qquad H = \tfrac{1}{2}g^{\alpha\beta}h_{\alpha\beta}.$$

With reference to any real number c, define the function $X^*(u^1, u^2)$ by (7). This vector function is of class C^1, since (6) is. It is readily verified

from (13)-(16) and (6) that the vector product of the partial derivatives $X^*_i(u^1, u^2)$ satisfies the identity

$$[X^*_1, X^*_2] = Q[X_1, X_2], \tag{17}$$

where $Q = Q^o = Q^o(u^1, u^2)$ is the (continuous, scalar) function

$$Q = 1 - 2cH + c^2K. \tag{18}$$

Let K_0, H_0 denote the values of K, H at any given point, $(u^1{}_0, u^2{}_0)$, of the (u^1, u^2)-domain and let, with reference to this point, the value of the arbitrary parameter c be chosen so as to be distinct from the roots of the quadratic equation

$$1 - 2cH_0 + c^2K_0 = 0 \tag{19}$$

(which becomes a linear equation or no equation at all according as $(u^1{}_0, u^2{}_0)$ is a non-umbilical or umbilical parabolic point of S, i. e., according as $K_0 = 0 \neq H_0$ or $K_0 = 0 = H_0$). Then the function (18), being distinct from 0 at the point $(u^1{}_0, u^2{}_0)$ of S, will be distinct from 0 in a vincinity of this point.

Let (u^1, u^2) be confined to such a vicinity. Then (17), where $[X_1, X_2] \neq 0$ by assumption, implies that $[X_1{}^*, X_2{}^*] \neq 0$. This means that (7) defines a surface $S^* = S^*(c)$ of class C^1. But (17), where $Q \neq 0$, also shows that the unit normal vector at a point (u^1, u^2) of S^* is identical with the unit normal vector, (6), of the S at the corresponding point (u^1, u^2) of S (with the understanding that, since (6) contains the ambiguity of a square root, N and $-N$ are not considered as distinct). Since (6) is a function of class C^1, it follows that S^* has a unit normal which is a function of class C^1 in the (u^1, u^2)-parametrization of S^*. It follows therefore by an application of the implicit function theorem that, if the plane tangent to S^* at $(u^1{}_0, u^2{}_0)$ is chosen to be the (x, y)-plane of the space $X = (x, y, z)$, then a (sufficiently small) vicinity of the point $(u^1{}_0, u^2{}_0)$ of S^* can be represented in the form $z = z(x, y)$, where $z(x, y)$ is a function of class C^2 (for details, cf. the corresponding argument in [3], p. 163). This proves (iii).

7. Denote by g^*_{ik}, h^*_{ik}, K^*, H^* the functions which result if $X(u^1, u^2)$ is replaced by $X^*(u^1, u^2)$ in the definitions (13), (14), (16), and refer by (j*), where $\mathrm{j} = 6, 13, \cdots$, to the formula which, on this replacement, results from formula (j). These definitions of the starred functions are valid even though $X^*(u^1, u^2)$ need not be of class C^2 in the (u^1, u^2)-parameters. However, it is clear from the tensor character of the matrices of the first

and second fundamental forms that K^* and H^* are actually the total and mean curvatures, respectively, of S^*, and that (g^*_{ik}) and (h^*_{ik}) are the matrices which result by a calculation of the first and second fundamental forms in a C^2-parametrization and then applying the standard transformation rule.[1]

It is readily verified from (15) and (13) that $N_i \cdot N_k = g^{\alpha\beta} h_{i\alpha} h_{k\beta}$. In terms of matrix multiplications, this can be written in the form

$$(N_i \cdot N_k) = (h_{ik})(g^{ik})(h_{ik}) \tag{20}$$

(incidentally, if $\alpha = (g^{ik})$ and $\beta = (h_{ik})$, and if a prime denotes transposition, then the matrix product (20) can be written as $\alpha\beta\alpha'$, since $\alpha' = \alpha$). Since $N^* = N$, it follows from (20) that

$$(h^*_{ik})(g^{*ik})(h^*_{ik}) = (h_{ik})(g^{ik})(h_{ik}). \tag{21}$$

The connection between the respective factors in (21) can be obtained as follows:

Since (7) implies that $X^*_i = X_i + cN_i$, it is seen from (13) and (13*) that g^*_{ik} is the sum of three terms, g_{ik}, $cX_i \cdot N_k + cX_k \cdot N_i$ and $c^2 N_i \cdot N_k$. In view of (14) and (15), this means that

$$(g^*_{ik}) = (g_{ik}) + 2c(h_{ik}) + c^2(h_{ik})(g^{ik})(h_{ik}) \tag{22}$$

and it is similarly verified, from (14) and (14*), that

$$(h^*_{ik}) = (h_{ik}) - c(h_{ik})(g^{ik})(h_{ik}). \tag{23}$$

Needless to say, the identity (21) could be verified from the relations (22) and (23). If the latter are combined with (16), (16*) and (18), there result the relations

$$K = QK^*, \qquad H = Q(H^* + cK^*). \tag{24}$$

While (24) is classical for a "smooth" pair S, S^* (in fact, for an S of class C^3, the corresponding S^* being of class C^2 for trivial reasons), the point in the following proof of (i) will be that (24) holds under the mere C^2-assumption of (iii).

8. Let S be a surface of C^2 and of constant positive curavture K.

---

[1] Needless to say, the matrix of the second fundamental form behaves under (u^1, u^2)-transformations (of class C^2) as a tensor only if the Jacobian of the transformation is positive, acquiring the factor -1 if the Jacobian is negative. This is precisely the reason why, in contrast to the case of the total curvature K, only the magnitude, and not the sign, of the mean curvature H has geometrical significance.

Then, with reference to any point $(u^1{}_0, u^2{}_0)$ of S, the quadratic equation (19) reduces to $Q_0 = 0$, where $Q_0 = Q(u^1{}_0, u^2{}_0; c)$ and, according to (18),

$$Q = Q(u^1, u^2; c) = 1 - 2cH + c^2, \qquad H = H(u^1, u^2),$$

if the value of the positive constant K is chosen to be 1. Since the product of the two roots of $Q_0 = 0$ is 1, either $c = 1$ or $c = -1$ is not a root (one of these c-values is a root if and only if the square of H_0 is 1, which, since $K_0 = K = 1$, means that $(u^1{}_0, u^2{}_0)$ is an umbilical point of S). It can be assumed that $c = 1$ is not a root, since, in view of the remark made after formula (2) above, H can be replaced by $-H$. Then $c = 1$ is not excepted in (iii), and so $S^* = S^*(1)$ is of class C^2.

The assumptions, $K = 1$ and $c = 1$, reduce (18) to $\frac{1}{2}Q = 1 - H$ (and $H = H(u^1, u^2)$ is distinct from 1 on S if S is a sufficiently small vicinity of its point $(u^1{}_0, u^2{}_0)$, since $Q_0 \neq 0$). Thus (24) reduces to the identities

$$1 = 2(1 - H)K^*, \qquad H = 2(1 - H)(H^* + K^*).$$

But substitution of the second of these identities into the first gives

$$H = 2(1 - H)H^* + 1, \text{ i.e., } (1 - H)(2H^* + 1) = 0$$

and (since $H \neq 1$ by the preceding parenthetical remark) the last equation implies that the function H^* of (u^1, u^2) must be the constant $-\frac{1}{2}$. This is the extension to the present case (of an S which is just of class C^2) of a classical result of Bonnet (cf., e. g., [2], p. 447).

9. The truth of (i) can now be concluded from (ii).

First, if S and H in (ii) are replaced by S^* and H^*, then, since S^* is of class C^2 and since H^*, being a constant, surely is of class $C^1(\lambda)$, it follows that S^* is of class C^3. Consequently, if S^* is written in the form $z = z^*(x, y)$ (locally), then $z^*(x, y)$ is a function of class C^3. Since $z^*(x, y)$ satisfies the differential equation which results if $H = H(x, y)$ in (2) is replaced by the constant $H^* = \frac{1}{2}$, it now follows from the general theorem of S. Bernstein (concerning elliptic differential equations; cf. Section 1) that the surface S^*: $z^* = z^*(x, y)$ is analytic. But the given S is a parallel surface of this S^*. Hence, the analyticity of S follows from that of S^*. This completes the proof of (i).

10. A corollary is the possibility of reducing to a geometrically significant minimum the smoothness assumptions in Liebmann's theorem concerning spheres, as follows:

If a closed, orientable surface is of (genus 0 and of) *class* C^2, *then* $K \equiv 1$ *cannot hold on it unless it is a sphere* (where, according to Hadamard [or just Gauss-Bonnet], the parenthetical restriction is superfluous).

Because of the use lines of curvatures in his partial differential equations, Hilbert's proof, as it stands, applies only under a C^n-restriction of fairly large n (something like $n = 5$); in this regard, cf. [3], pp. 163-172. But (i) reduces this n to $n = 2$ (S. Bernstein's reduction reaches to $n = 3$).

Appendix.

On tangential representations.

Let the surface (5) or S: $z = z(x, y)$ be of class C^2, sufficiently small and free of parabolic points (i. e., let either $K > 0$ or $K < 0$ on S). Then N, being defined by (6), is a vector function of class C^1, and the (local) theorem on implicit functions shows that the correspondance between S: $X = (x, y, z)$ and its normal image N: (ξ, η, ζ) is one-to-one and such that

$$S\colon X = X(N) \tag{25}$$

is a vector function of class C^1 (on the corresponding portion of the unit sphere $|N| = 1$). Hence the same is true of the scalar function $p = N \cdot X$ (Weingarten; cf. [2], pp. 135-137) and therefore of its homogenized form [2]

$$H = \rho p(\xi/\rho, \eta/\rho, \zeta/\rho), \text{ where } \rho = (\xi^2 + \eta^2 + \zeta^2)^{\frac{1}{2}} > 0, \qquad N = (\xi/\rho, \eta/\rho, \zeta/\rho). \tag{26}$$

Thus, if S: $X(u^1, u^2)$ is of class C^2, sufficiently small and of non-vanishing $K = K(u^1, u^2)$, then $H = H(\xi, \eta, \zeta)$ is a function of class C^1 (within that infinite cone of the (ξ, η, ζ)-space which corresponds to S in the spherical mapping of Gauss). But it turns out that, under the assumptions just stated, *the function* $H = H(\xi, \eta, \zeta)$, *instead of being only of class* C^1, *is of class* C^2 *by necessity.*

This lemma, which in a certain sense is an analogue of (iii) in Section 4 above, is unexpected, since it implies that, if parabolic points are excluded,

[2] This H, which is Minkowski's notation for the supporting function (26) (if $K < 0$ is excluded and S is closed and orientable), will not be confused with the preceding H, which is the mean curvature.

then Minkowski's applications of his *general* theory to *differential* geometry hold under the true assumption of the differential geometry of surfaces, which is the assumption of surfaces of class C^2. In contrast, the literature consulted says or implies that Minkowski's theory is applicable only under the artificial assumption that the surfaces considered are confined to the class C^3. Suffice it to say that the mean curvature and the Gaussian curvature are expressible in terms of the second derivatives of H provided that H is of class C^2, while the argument applied after (25) assures the C^2-character of H (i. e., of the representation of S in *plane coordinates*) only if (5) (i. e., the representation of S in terms of *point coordinates*) is of class C^3.

Incidentally, it will be clear from the proof below that the last italicized assertion can be generalized, to every $n \geqq 2$, in the same way as (iii*) in Section 5 generalizes (iii) in Section 4. In the excluded case, $n = 1$, the assertion is now not only false but meaningless as well, since the curvature $K = K(u^1, u^2)$ of a surface of class C^n cannot in general be defined if $n = 1$. On the other hand, it is clear (even in the case of an analytic S) that the assumption $K \neq 0$ cannot be omitted.

The proof (for $n = 2$) of the last italicized assertion proceeds as follows:

Since the piece of surface (5) is supposed to be sufficiently small and of class C^2, it can be assumed that S is given in the form $S\colon z = z(x, y)$, where (x, y) varies on a small domain, say the circle $x^2 + y^2 < \epsilon^2$, on which the function $z(x, y)$ is of class C^2. It can also be assumed that the oriented unit normal $N = (\xi, \eta, \zeta)$ of S at $(x, y) = (0, 0)$ is in the positive half of the z-axis, hence $\zeta > 0$ on $x^2 + y^2 < \epsilon^2$ (if ϵ is small enough). Then, if

$$(27) \qquad q = \xi/\zeta, \qquad r = \eta/\zeta,$$

it is seen from (26), where $p = X \cdot N$ and $X = (x, y, z)$, that

$$(28) \qquad qx + ry = H + L,$$

where

$$(29) \qquad H = H(q, r, 1) \text{ and } L = -z = -z(x, y) \equiv L(x, y)$$

are functions of class C^1 and C^2, respectively. But (28) and (29) imply that

$$(30) \qquad x = H_q, \qquad y = H_r,$$

which, together with (28), defines a Legendre transformation.

In fact, in view of the involutory character of a Legendre transformation, (30) and (28) together are equivalent to

$$(31) \qquad q = L_x, \qquad r = L_y$$

and (28) together, provided that $L = L(x, y)$ is a function of class C^2 and of non-vanishing Hessian (i. e., provided that the transformation (31) of (x, y) into (q, r) is of class C^1 and of non-vanishing Jacobian). But the definition (29) of L shows that this proviso is satisfied, since the function $z(x, y)$ is of class C^2 and has a Hessian which, being a positive multiple of the curvature $K = K(x, y)$, does not vanish by assumption.

It now follows from a known fact on Legendre transformations (cf. [7], p. 6) that the transformation (30) of (q, r) into (x, y) is of class C^1 (and of non-vanishing Jacobian). But this means that $H = H(q, r, 1)$ in (30) is a function of class C^2 in (q, r). It follows therefore from (27) and (26) that $H(\xi, \eta, \zeta)$ is a function of class C^2 in $(\xi, \eta, \zeta) \neq (0, 0, 0)$.

From the formal point of view, the supporting function of a parallel surface results by adding a constant to the supporting function of the given surface. It is clear that, because of this formal circumstance, the fact just verified could be applied in the direction of (iii).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] S. Bernstein, "Sur la déformation des surfaces," *Mathematische Annalen*, vol. 60 (1905), pp. 434-436.

[2] L. Bianchi, *Lezioni di geometria differenziale*, Piza, 1894.

[3] P. Hartman and A. Wintner, "On the asymptotic curves of a surface," *American Journal of Mathematics*, vol. 73 (1951), pp. 149-172.

[4] L. Lichtenstein, "Ueber den analytischen Charakter der Lösungen regulärer zweidimensionaler Variationsprobleme," *Bulletin International de l'Académie des Sciences de Cracovie, Classe des Sciences Mathématiques et Naturelles*, sér. A, 1913, pp. 915-941.

[5] J. Steiner, *Gesammelte Werke*, vol. 2 (1882), pp. 170-176.

[6] A. Wintner, "On isometric surfaces," *American Journal of Mathematics*, vol. 74 (1952), pp. 192-214.

[7] ———, *The Analytical Foundations of Celestial Mechanics*, Princeton, 1941.

THE SPECTRA OF QUANTUM-MECHANICAL OPERATORS.*

By C. R. Putnam.

1. Let A and B denote "Hermitian" operators (in a sense to be defined), satisfying the commutator "equation"

$$AB - BA = iI, \tag{1}$$

where, here, and throughout this paper, i is the imaginary unit, I denotes the identity operator, and all operators are to be regarded with respect to an underlying Hilbert space $\mathfrak{H}$. The justification for the appearance of the quotation marks above is that the unqualified relation (1) can *never* hold as an actual equation. Specifically, the identity operator appearing on the right side of (1) is defined for all elements f in $\mathfrak{H}$; thus, if (1) were a true equation, then clearly both A and B would be defined for all f in $\mathfrak{H}$ and hence would be bounded operators,[1] in contradiction to known results; cf. [12], [8], [10]. What is needed, then, is the specification of a domain, say Ω, of elements in $\mathfrak{H}$ for which (1) does hold, that is for which

$$(AB - BA)f = if, \tag{2}$$

for all f in Ω. The statement that "(1) holds on Ω" will then be taken to mean that (2) holds for all f in Ω. It should be noted that Ω belongs both to the domain of definition of the operator AB and to that of BA.

This paper will be devoted to the problem of determining the nature of A and B and their spectra in relation to the domain Ω. Certain results in this direction have recently been obtained by Rellich [8] under assumptions quite different to those of the present paper. In fact, the conditions imposed in [8] are such as to imply that the operators A and B are of the Heisenberg-Schroedinger type, and are therefore both unbounded; whereas in the case at hand, the fact that the boundedness of one of the operators is not precluded is essentially connected with the nature of the theorems to be presented below. For other results bearing on problems somewhat related to those considered in the present paper, see [7].

* Received March 3, 1951.

[1] If A is any symmetric operator for which $\mathfrak{D}_A = \mathfrak{H}$ (cf. section 2) then necessarily A is self-adjoint and is bounded ([9], pp. 51, 59).

The significance of the role of Ω in determining the spectral properties of A and B can perhaps be most readily demonstrated if A and B are defined by

$$A = id/dx, \qquad B = x. \tag{3}$$

If f denotes any differentiable function, then clearly (2) holds, so that in a sense, (1) can be regarded as a "formal" identity. First, suppose that $\mathfrak{H}$ consists of all (complex-valued) functions f of class $L^2(-\infty, \infty)$. Then (1) holds on a certain subset Ω of $\mathfrak{H}$, and furthermore it is known ([5], pp. 66 ff.) that A and B can be regarded as self-adjoint operators (as defined in the next section), and that each posssesses the whole line $-\infty < \lambda < \infty$ as its spectrum; in addition, neither A nor B possesses an eigenvalue.

Next, let $\mathfrak{H}$ consist of all functions of class $L^2[a, b]$, where $[a, b]$ denotes a *finite* interval. If the boundary condition $|f(a)/f(b)| = 1$ (among other conditions) is required for the functions in the domain of A, it is known that again both A and B can be regarded as self-adjoint operators. In this case, however, the spectrum of A consists of a discrete sequence of eigenvalues $\lambda_1, \lambda_2, \cdots$, which satisfy $\lambda_n \to \infty$ as $n \to \infty$, while the spectrum of B is the entire interval $a \leq \lambda \leq b$.

It was thought by Dirac ([1], p. 94) that the situation in which the spectra of A and B consist of every real number was a consequence of the "equation" (1). However, as shown by the last example (which has been cited in a quantum-mechanical connection in the physical literature; [2], p. 299) such is, in general, not the case. Moreover, as is shown in Appendix II below, it is quite possible that A and B be formally Hermitian matrices (so that $a_{jk} = \bar{a}_{kj}$ and $b_{jk} = \bar{b}_{kj}$) which satisfy (1), and are such that one of the pair is bounded (and hence possesses a spectrum contained in a finite interval). In fact, the matrix consideration referred to here reflects the situation for the operators (3) on the finite interval.

As previously remarked, the present paper will deal with the problem of characterizing the operators A and B (both of which, except in Theorem I and the Lemma below, are self-adjoint) satisfying (1), in regard to Ω. A different type of problem, concerning the nature of the spectrum of a certain self-adjoint differential equation and the concomitant boundary value, problem, has been considered by Wintner [13].

The discussion of the cases bearing on (3) considered above, will be considered further in section 6.

2. Terminology. The underlying Hilbert space will be denoted by $\mathfrak{H}$;

it is understood that $\mathfrak{H}$ is not the trivial space consisting of the zero element alone. The definitions and terminology will be that of Nagy [3]. Thus, an operator A will be called symmetric if A is linear and densely defined, and if $(Af, g) = (f, Ag)$ for all elements f, g in $\mathfrak{D}_A$, the domain of A; hence $A \subseteqq A^*$. An operator A will be called normal if A is linear, densely defined, closed, and satisfies $AA^* = A^*A$; in particular, A will be called self-adjoint if A is normal and $A = A^*$. It is to be emphasized that the *unqualified* equation $A = B$ entails both relations $A \subseteqq B$ and $A \supseteqq B$.

By $\mathfrak{R}_A$ will be meant the range of the operator A, so that $\mathfrak{R}_A$ consists of all elements g in $\mathfrak{H}$ such that $g = Af$, f in $\mathfrak{D}_A$. Finally, if Ω denotes any set contained in $\mathfrak{D}_A$, then $A(\Omega)$ denotes the image of Ω, that is, the set of all elements g in $\mathfrak{H}$ such that $g = Af$, f in Ω.

All operators considered in this paper are linear.

3. The theorems and some of the proofs will be given in this section and in section 5; the remaining proofs (those of Theorems I, II and III) will be given in sections 7-10.

THEOREM I. *Let A be symmetric and let B be an arbitrary linear operator in $\mathfrak{H}$ satisfying the commutator equation* (1) *on a set Ω of $\mathfrak{H}$. Then A cannot possess a real eigenvalue λ with an eigenfunction belonging to Ω. Moreover, the above statement remains true if A and B are interchanged.*

It is to be noticed that the above theorem does not preclude the existence of eigenfunctions for the operators A or B; in fact, it is quite possible that both A and B are self-adjoint, satisfy (1) on a set Ω, and that the entire spectrum of A consists of (real) eigenvalues (with corresponding eigenfunctions which, by Theorem I, cannot belong to Ω). Such an example is furnished by (3) with reference to the finite interval $[a, b]$; cf. the remarks in section 1. The situation is readily explained in the present instance by the fact that if a function $f(x)$ is in $\mathfrak{D}_A$, it satisfies the boundary condition $|f(a)/f(b)| = 1$. Since, however, $\Omega \subseteqq \mathfrak{D}_{AB}$, all functions of Ω necessarily satisfy this same condition; in particular if f is an eigenfunction of A belonging to Ω, then both $f(x)$ and $xf(x)$ would have to satisfy the boundary condition. This is impossible since $|f(a)| = |f(b)| \neq 0$ for such f's.

Theorem I is similar to a portion of Theorem III of [7].

The next theorem is

THEOREM II.[2] *Let A and B denote self-adjoint operators in $\mathfrak{H}$ satisfying*

[2] The author wishes to thank Professor Kodaira for several valuable comments and suggestions in connection with the formulation and proof of this theorem.

(1) *on a set* Ω, *and suppose that the set* $AB(\Omega)$ *is dense in* $\mathfrak{H}$. *In addition, let* A^{-1} *be bounded. Then* (i) $\mathfrak{D}_A \subseteqq \mathfrak{R}_B$ *and* (ii) A *is unbounded. Moreover, the theorem remains true if* A *and* B *are everywhere interchanged.*

The next theorem is similar to the above but with a slightly altered hypothesis.

THEOREM III. *The assertions of Theorem II above remain valid if the assumption that* $AB(\Omega)$ *be dense in* $\mathfrak{H}$ *is replaced by the assumption that* $\Phi = BA(\Omega)R_{AB}$ *be dense in* $\mathfrak{H}$.

In the next section, a lemma will be proved which will be useful in deriving certain consequences of the above theorems.

4. For an arbitrary pair of complex numbers λ, μ and an arbitrary pair of operators A, B, define A_λ and B_μ by

$$A_\lambda = A - \lambda I, \qquad B_\mu = B - \mu I. \tag{4}$$

It is readily seen that if λ denotes a real number, then the operator A_λ is self-adjoint (symmetric) if and only if A is self-adjoint (symmetric). First, the following Lemma will be derived:

LEMMA. *If* A *and* B *denote arbitrary operators (not necessarily self-adjoint or even symmetric) in a Hilbert space* $\mathfrak{H}$, *and if* (1) *holds on* Ω, *then*

$$A_\lambda B_\mu - B_\mu A_\lambda = iI \qquad (\lambda \text{ and } \mu, \text{ arbitrary complex numbers}), \tag{5}$$

also holds on Ω.

Proof. Since $\Omega \subseteqq \mathfrak{D}_{AB} \subseteqq \mathfrak{D}_B$, and $\Omega \subseteqq \mathfrak{D}_{BA} \subseteqq \mathfrak{D}_A$, it is clear from (1) that

$$AB - \mu A - \lambda B + \lambda\mu I - BA + \mu A + \lambda B - \lambda\mu I = iI \tag{6}$$

holds on Ω. Since $AB + AC \subseteqq A(B + C)$, and $(A + B)C = AC + BC$ hold for arbitrary operators A, B, C (cf. [3], p. 28), relation (6) implies first that

$$A(B - \mu I) - \lambda(B - \mu I) - (B - \mu I)A + \lambda(B - \mu I) = iI$$

holds on Ω, and then that

$$(A - \lambda I)(B - \mu I) - (B - \mu I)(A - \lambda I) = iI$$

holds on Ω. As this last relation is simply (5), the proof of the Lemma is complete.

5. It is clear that if the role of (1) in Theorems I-III is played by (5), and if the role of the set Ω is played by some set $\Omega_{\lambda\mu}$, then theorems similar to I-III, but pertaining to A_λ and B_μ, where λ and μ are real, may readily be obtained. According to the Lemma, the set $\Omega_{\lambda\mu}$, on which (5) holds, contains the set on which (1) holds, that is, $\Omega \subseteqq \Omega_{\lambda\mu}$. Since another application of the Lemma shows that $\Omega_{\lambda\mu} \subseteqq (\Omega_{\lambda\mu})_{-\lambda-\mu}$ $(=\Omega)$, it follows, however, that even $\Omega = \Omega_{\lambda\mu}$ (λ, μ complex, arbitrary) is valid.

In addition it can be remarked that in order to prove, for instance, that A is unbounded, it is sufficient to prove that A_λ is unbounded for some (real or complex) value λ.

For the sake of application to the problem of locating spectra, the following theorem, which involves A_λ and B_μ, and which is a corollary of Theorem II, will be stated.

THEOREM IV. *Let A and B denote self-adjoint operators in $\mathfrak{H}$ satisfying* (1) *on a set Ω, and let A_λ and B_μ be defined by* (4).

(i) *Consider the set, S, of real values λ for which there is a real value $\mu = \mu(\lambda)$ such that the set $A_\lambda B_\mu(\Omega)$ is dense in $\mathfrak{H}$ and $\mathfrak{D}_{A_\lambda}$ is not contained in $\mathfrak{R}_{B_\lambda}$. Then the set S is contained in the spectrum of the operator A.*

(ii) *The above statement* (i) *remains true if the symbols A and B are interchanged throughout.*

Remark. It should be noted that if a self-adjoint operator A possesses the spectral resolution $A = \int_{-\infty}^{\infty} \alpha dE_\alpha$, then $A_\lambda = \int_{-\infty}^{\infty} (\alpha - \lambda) dE_\alpha$. Furthermore, if λ is not an eigenvalue of A, then the inverse of A_λ, that is A_λ^{-1}, exists and is given by $A_\lambda^{-1} = \int_{-\infty}^{\infty} (\alpha - \lambda)^{-1} dE_\alpha$, the resolvent of A. Since A is a self-adjoint (and therefore normal) operator, it can possess at most a denumerable set of eigenvalues λ (for which values A_λ^{-1}, of course, fails to exist). A point λ is *not* in the spectrum of A if and only if A_λ^{-1} exists and is bounded. In addition, it should be noted that if λ is not an eigenvalue, then the inverse operator A_λ^{-1} (as well as A_λ) as defined above is self-adjoint whenever λ is real.

In connection with the above remark see [3], p. 54.

Proof of Theorem IV. Since an interchange of A and B in (1) is equivalent to the substitution of $-A$ for A in this relation, it is clearly sufficient to prove only part (i) of Theorem IV. By virtue of the Lemma, relation

(5) holds on Ω. Let λ be in S, and suppose λ is not in the spectrum of A. Then A_λ^{-1} exists, is self-adjoint, and is bounded. An application of (i) of Theorem II then implies $\mathfrak{D}_{A_\lambda} \subseteqq \mathfrak{R}_{B_\mu}$, in contradiction to the hypothesis; hence all points of S are in the spectrum of A. This completes the proof of Theorem IV.

6. Examples. Consider the real axis $-\infty < x < \infty$, and on it the self-adjoint operators A and B defined by (3) (cf. section 1). It is known ([5], pp. 66 ff.) that both A and B possess the set $-\infty < \lambda < \infty$ as their spectra. A proof of this fact, as far as A is concerned, will be given as an application of Theorem IV. Since A is self-adjoint, the spectrum of A is contained on the real line $-\infty < \lambda < \infty$. It will be shown that for any real λ, the function $\mu = \mu(\lambda) \equiv 0$ suffices in part (i) of Theorem IV. Thus, it will first be shown that the set $A_\lambda B(\Omega)$ is dense for all real λ. (It is to be noted that $B_0 = B$.) This set consists of all functions g such that

$$g = if' - \lambda f, \tag{7}$$

where $f = xF$, and F is in Ω. Multiplication of (7) by $e^{i\lambda x}$ yields

$$ge^{i\lambda x} = i(fe^{i\lambda x})'. \tag{8}$$

Let $k = k(x)$ be in $L^2(-\infty, \infty)$ and satisfy

$$\int_{-\infty}^{\infty} kg\, dx = 0 \text{ for all functions } g \text{ of (7).} \tag{9}$$

Consider only functions F with continuous first derivatives on $-\infty < x < \infty$ and which, in addition, are such that $f \equiv 0$ outside of a finite interval $[a, b]$. Clearly the functions F, as well as the functions $f = xF$, belong to the class Ω and are dense in $\mathfrak{H}$. From $\int_a^b kg\, dx = 0$, one obtains $i \int_a^b ke^{-i\lambda x}(fe^{i\lambda x})' dx = 0$. Put $K(x) = ike^{-i\lambda x}$; then the last relation implies $\int_a^b fe^{i\lambda x} dK(x) = 0$ for all functions f of the type considered above. This clearly implies that $K = $ const. on $[a, b]$. Since a and b are arbitrary, $K = $ const. on $-\infty < x < \infty$. Since K belongs to class $L^2(-\infty, \infty)$, K (and hence k) is identically zero. Thus the set $A_\lambda B(\Omega)$ is dense for *every* λ. The assertions regarding A will follow if it is shown that $\mathfrak{D}_{A_\lambda}$ is not contained in $\mathfrak{R}_B$ for any real λ. However, if a function f is in $\mathfrak{R}_B$, then $f = xg$, where $g \subset \mathfrak{D}_B$, and consequently $f/x \subset L^2(-\infty, \infty)$. But the existence of functions f in $\mathfrak{D}_{A_\lambda}$ for which f/x is not in $L^2(-\infty, \infty)$ is clear; in fact, one need only arrange that $f(0) \neq 0$,

so that (f being continuous), $\int_{-\infty}^{\infty} | f/x |^2 dx = \infty$ holds. Consequently, the spectrum of A contains the entire line $-\infty < \lambda < \infty$, and is therefore just this line.

The situation for the line $-\infty < x < \infty$ is to be contrasted with that for the finite interval $[a, b]$. In the latter case, as remarked in section 1, the boundary condition $| f(a)/f(b) | = 1$; (among other requirements) must be imposed on the functions of $\mathfrak{D}_A$. Since, among other things, all functions of Ω must satisfy the boundary condition $f(a) = f(b) = 0$, one can readily see, by direct verification, that the set $\Gamma = A_\lambda B_\mu(\Omega)$ is, in the present case, never dense (λ, μ real, arbitrary), and that the set $\Phi = B_\mu A_\lambda(\Omega)$ is surely not dense if B_μ^{-1} is bounded. In fact, the function $k = e^{i\lambda x}$ is orthogonal to all elements of Γ; while if B_μ^{-1} is bounded, so that μ does not belong to $[a, b]$, then the function $k = e^{i\lambda x}(x - \mu)^{-1}$ is orthogonal to all elements of Φ.

7. *Proof of Theorem* I. It is sufficient to prove the theorem in the case that A is symmetric and B is arbitrary; cf. the above remark at the beginning of the proof of Theorem IV. According to the Lemma, (5) holds on Ω (λ and μ arbitrary). Hence, if A possesses a real eigenvalue with an eigenfunction f in Ω, then $A_\lambda f = 0$ for some real λ, and if $\mu = 0$, then (5) implies $A_\lambda Bf - BA_\lambda f = if$, so that, since B is linear, $A_\lambda Bf = if \subset \Omega$. In particular, since $\Omega \subseteqq \mathfrak{D}_{BA_\lambda} \subseteqq \mathfrak{D}_{A_\lambda}$, the element $A_\lambda Bf \subset \mathfrak{D}_{A_\lambda}$. Since λ is real, A_λ is symmetric, and hence $(g, A_\lambda h) = (A_\lambda g, h)$ whenever g, h belong to $\mathfrak{D}_{A_\lambda}$. Clearly $Bf \subset \mathfrak{D}_{A_\lambda}$. Let $g = A_\lambda Bf$ and $h = Bf$; then $(A_\lambda Bf, A_\lambda Bf) = (A_\lambda A_\lambda Bf, Bf) = 0$, since $A_\lambda A_\lambda Bf = A_\lambda(if) = 0$. Thus $A_\lambda Bf = 0$ and hence $f = 0$, in contradiction to the requirement that f be an eigenfunction. This completes the proof of Theorem I.

8. *Proof of* (i) *of Theorem* II. As above, it is sufficient to prove the theorem as stated in terms of A and B (cf. the beginning of the proof of Theorem IV above). It follows from (1) that $2\mathfrak{J}(ABf, f) = (f, f)$ so that

$$\| ABf \| \geq \tfrac{1}{2} \| f \|, \qquad f \text{ in } \Omega. \tag{10}$$

In order to prove (i), it is sufficient to show that for any element g in $\mathfrak{H}$, there exists an element f of $\mathfrak{D}_B$ for which $A^{-1}g = Bf$. Since $AB(\Omega)$ is dense in $\mathfrak{H}$, there exists a sequence $f_1, f_2, \cdots$, where f_n are in Ω, such that $ABf_n \to g$ as $n \to \infty$. By virtue of (10), $f_n \to f \subset \mathfrak{H}$, as $n \to \infty$. But A^{-1} is bounded, so that $Bf_n \to A^{-1}g$; since B is closed, f is in $\mathfrak{D}_B$, and $Bf = A^{-1}g$. This proves (i).

9. *Proof of* (ii) *of Theorem* II. It will be shown that the assumption that A is bounded leads to a contradiction. Suppose, if possible, that A (as well as A^{-1}) is bounded. Since $\mathfrak{D}_A = \mathfrak{H}$ and, by (i), $\mathfrak{D}_A \subseteqq \mathfrak{R}_B$, it follows that B^{-1} exists and is self-adjoint (cf. [3], p. 35, (c)). In addition, B^{-1} is also bounded ([9], p. 59. See also the above footnote 1). Let f belong to $AB(\Omega)$, so that $f = ABg$, where $g \subset \Omega$. An application of (1) to $g = B^{-1}A^{-1}f$ yields $f - BAB^{-1}A^{-1}f = iB^{-1}A^{-1}f$. Let α and β denote the bounded operators $A^{-1}B^{-1}$ and $B^{-1}A^{-1}$ respectively. Then an application of the operator α to the last equation gives

$$\alpha f - \beta f = i\alpha\beta f, \tag{11}$$

which is valid for all f in $AB(\Omega)$ (and hence for all f in $\mathfrak{H}$). Clearly $\alpha = \beta^*$; therefore if U is defined by

$$U = \beta + iI, \tag{12}$$

relation (11) implies

$$U^*U = I. \tag{13}$$

For any bounded operator C, let $\| C \|$ be defined by $\| C \| = \text{l. u. b.} \| Cf \|$ when $\| f \| = 1$; since $(Uf, Uf) = (f, U^*Uf) = \| f \|^2$ for all $f \subset \mathfrak{H}$, it follows that $\| U \| = 1$ (and hence $\| U^* \| = 1$). Moreover, the spectrum of any bounded operator C is contained in the circle $| z | \leqq \| C \|$ of the complex plane (cf. [11], p. 143). Hence, by (12), the spectra of β and β^* lie in circles of radii 1 with centers at $z = -i$ and at $z = i$, respectively.

Since $A^{-1}B^{-1} = A^{-1}(B^{-1}A^{-1})A$, it follows from the fact that A and A^{-1} are bounded that β and β^* $(= \alpha)$ have identical spectra (cf. [12]). Consequently, the origin $z = 0$ is the only possible point in the spectra of β and β^*. (Incidentally, that $z = 0$ actually is in the spectrum of β is clear from the fact that the spectrum of a bounded operator is never empty; cf. [11], pp. 143-144.) Hence by (12), $\lambda = 0$ certainly is not in the spectrum of U, so that U is non-singular (possesses a unique, two-sided, bounded inverse); since, as was seen above, $U^*U = I$, it follows that $UU^* = I$ must also hold. Thus U is unitary and β is normal, so that $\beta = B^{-1}A^{-1}$ is the (bounded) zero operator. (It should be noted that the fact that $z = 0$ is the only point in spectrum of a *non-normal* bounded operator C is *not* sufficient to guarantee that C is the zero operator.) It follows that β possesses $\lambda = 0$ as an eigenvalue with any $f \neq 0$ in $\mathfrak{H}$ as a corresponding eigenfunction. One thereby obtains a contradiction to the fact that β is the inverse of AB (cf. [3], p. 28). In fact, $0 = \beta ABf = f$ for all f in $\mathfrak{D}_{AB}$; the existence of some $f \neq 0$ in $\mathfrak{D}_{AB}$, however, follows from the assumption that $AB(\Omega)$ is dense in the (non-trivial) space $\mathfrak{H}$. This contradiction completes the proof of Theorem II.

10. *Proof of Theorem* III. Since $\mathfrak{R}_B$ is dense, B^{-1} exists and is self-adjoint; [3], p. 35, (c). Let $f \subset \Phi$; then $f = BAg$, $g \subset \Omega$. An application of (1) to $g = A^{-1}B^{-1}f$ yields $ABA^{-1}B^{-1}f - f = iA^{-1}B^{-1}f$. As in section 9, let $\alpha = A^{-1}B^{-1}$ and $\beta = A^{-1}B^{-1}$ (with the understanding that α and β need not be bounded in the present case). Since f is in Φ, both elements f and $ABA^{-1}B^{-1}f$ belong to $\mathfrak{D}_\beta$, so that $A^{-1}B^{-1}f$ must also belong to $\mathfrak{D}_\beta$. An application of β to the above equation gives

$$\alpha f - \beta f = i\beta\alpha f, \qquad f \text{ in } \Phi.$$

Since Φ is dense, both α and β are densely defined and, consequently, the adjoints α^* and β^* exist. Since A^{-1} is bounded, $\alpha^* = \beta$ ([3], top of p. 29). Hence β, being an adjoint, is closed. In addition, $\alpha \subseteqq \beta^*$; [3], top of p. 29. The last formula line now implies $\beta^* f - \beta f = i\beta\beta^* f$, for all f in Φ. If, as in section 9, U is defined by (12), it is clear that U is linear, closed, and densely defined, that U^* exists, and that (cf. [3], top of p. 29) $U^* = \beta^* - iI$. It follows that

$$UU^* f = f, \qquad f \text{ in } \Phi. \tag{14}$$

Furthermore $U^{**} = U$ ([3], middle of p. 29). Since (14) holds for all f in Φ, $\Phi \subseteqq \mathfrak{D}_{U^{**}U^*} \subseteqq \mathfrak{D}_{U^*}$. Let h be an arbitrary element in $\mathfrak{H}$; since Φ is dense, there exists a sequence of elements f_n in Φ such that $f_n \to h$ as $n \to \infty$. Clearly

$$(U^*(f_n - f_m), U^*(f_n - f_m)) = ((f_n - f_m), U^{**}U^*(f_n - f_m)),$$

so that by (14) and the equation $U^{**} = U$, $U^*(f_n - f_m) \to 0$. Thus $U^* f_n \to k$, where k is some element of $\mathfrak{H}$. But U^*, being an adjoint, is closed; hence, $U^* h = k$, that is, $U^* h$ is defined for all h in $\mathfrak{H}$. Furthermore, it is readily seen that $(U^* f, U^* f) = (f, U^{**}U^* f) = (f, f)$ for f in Φ. Thus $|U^* f| = |f|$ whenever f is in Φ, and hence (by the above results) whenever f is in $\mathfrak{H}$. Consequently, U^* is bounded; this implies that $U (= U^{**})$ is bounded. It now follows from (12) that β is bounded, and therefore $\mathfrak{R}_{A^{-1}} \subseteqq \mathfrak{D}_{B^{-1}}$, that is, $\mathfrak{D}_A \subseteqq \mathfrak{R}_B$. This proves the first part of Theorem II. The proof of the second part, namely that A is unbounded, is similar to that given in the latter part of section 9. Thus, the assumption that A is bounded readily implies that $U^*U = I$ as well as $UU^* = I$ (the last relation having been obtained above), so that U is unitary. The existence of an $f \neq 0$ in $\mathfrak{D}_{AB}$ follows from the fact that $\Phi \subseteqq \mathfrak{R}_{AB}$, and one obtains a contradiction as before. This completes the proof of Theorem III.

Appendix.

I. Certain formulations of the statement "(1) holds on Ω" can even specify explicitly the set Ω. For instance, if the assumption that (1) holds on Ω is replaced by $AB \subseteqq BA + iI$, then $\Omega = \mathfrak{D}_{AB}$. Furthermore, if it is assumed that $AB = BA + iI$, then $\Omega = \mathfrak{D}_{AB} = \mathfrak{D}_{BA}$. Thus under such revised assumptions, the formulations of the various theorems of this paper can be stated more simply; these revised forms are clear, however, and will not be stated explicitly.

Also, slight variations of the above theorems and even sharpened statements can be obtained in certain instances. Such a case in point is offered by part (ii) of Theorem II. In fact, according to Theorem III of [7], if A is arbitrary and B is normal (not necessarily self-adjoint), then the relation $AB \subseteqq BA + iI$ implies, without further assumptions, that A is unbounded.

II. The existence of a pair of matrices $A = (a_{mn}), B = (b_{mn})$ satisfying relation (1), the Hermitian requirements

$$(15) \qquad (a_{mn}) = (\bar{a}_{nm}), \qquad (b_{mn}) = (\bar{b}_{nm}),$$

and

$$(16) \qquad \sum_m |a_{mn}|^2 < \infty, \qquad \sum_m |b_{mn}|^2 < \infty, \qquad n = 1, 2, \cdots,$$

and such that one of the pair is bounded, will be shown. According to (16), the rows (hence, by (15), the columns) of A and B belong to Hilbert space; as a consequence, the products AB and BA occurring in (1) (the latter relation being now regarded as an ordinary matrix equation) surely exist. Specifically it will be shown that the pair A, B defined by

$$(17) \qquad a_{mn} = \begin{cases} 4imn/\pi(m^2 - n^2), \ m \text{ odd (even) and } n \text{ even (odd)} \\ 0, \text{ otherwise} \end{cases}$$
$$b_{mn} = \begin{cases} \pi/2, \ m = n \\ -8mn/\pi(m^2 - n^2)^2, \ m \text{ odd (even) and } n \text{ even (odd)} \\ 0, \text{ otherwise} \end{cases}$$

satisfies (1), (15) and (16), and that B is bounded. In principle, the verification of these claims can be made directly; however, the proof, which will be given below, will depend on a correspondence between matrices and operators.

Consider the operators $A = id/dx$ and $B = x$, which satisfy the "formal" identity

$$(18) \qquad AB - BA = i,$$

as operating on the class, $\mathfrak{F}$, of functions $f(x)$, which are absolutely continuous on the *finite* interval $a \leqq x \leqq b$ and satisfy

$$f(a) = f(b) = 0, \qquad \int_a^b |f'|^2 dx < \infty \qquad (' = d/dx). \tag{19}$$

It is clear that the set $\mathfrak{F}$ is dense in the Hilbert space $\mathfrak{H}$ of all functions f on $[a, b]$ for which $\int_a^b |f|^2 dx < \infty$, and that, as is seen from (18) and (19),

$$\int_a^b f(\overline{Cg})\, dx = \int_a^b (Cf)\bar{g}\, dx, \text{ where } C = A \text{ or } B \text{ and } f, g \text{ are in } \mathfrak{F}$$

(cf. [5], pp. 66, 68, and [8], p. 111). If desired, the domain of definition of the (unbounded) operator A can be extended so that A becomes a closed operator (cf. [5], pp. 76, 79); it should be noted that B is a bounded operator (and is defined for all f in $\mathfrak{H}$). If $\phi_1(x), \phi_2(x), \cdots$ denotes any orthonormal complete sequence of functions of class $\mathfrak{F}$, and if the matrices $A = (a_{mn})$ and $B = (b_{mn})$ are defined by

$$a_{mn} = i \int_a^b \phi_m \bar{\phi}_n' dx, \qquad b_{mn} = \int_a^b x \phi_m \bar{\phi}_n dx,$$

then A and B satisfy (15) and (16), and B is bounded (cf. [11], pp. 184 ff., and [4], pp. 122 ff.). In addition, as a consequence of (18) and (19), it is seen by direct verification, that (1) also holds. By choosing $[a, b]$ to be $[0, \pi]$ and $\phi_n(x) = (2/\pi)^{\frac{1}{2}} \sin nx$, one obtains the matrices A and B of (17). Incidentally, it can be remarked that the spectrum of the matrix B defined by (17) is exactly the interval $[0, \pi]$ ([11], pp. 184 ff.).

By choosing different sequences $\phi_1(x), \phi_2(x), \cdots$ on $[a, b]$, one can obtain various pairs A, B satisfying (1), (15) and (16), and such that B is bounded. All of the bounded matrices B obtained in this manner, are unitarily equivalent to each other ([11], pp. 184 ff. and [4], pp. 122 ff.) and none, of course, is unitarily equivalent to either of the linear oscillator matrices occurring in physics (cf., e. g., [2], p. 370). It will remain undecided whether *all* bounded matrices B, which have identical spectra and which can occur in pairs A, B satisfying (1), (15) and (16), are unitarily equivalent to each other, or what amounts to the same thing, to a fixed such matrix, say the one defined by the second relation of (17), in case the spectrum is the interval $[0, \pi]$.

THE INSTITUTE FOR ADVANCED STUDY.

REFERENCES.

[1] P. A. M. Dirac, *The Principles of Quantum Mechanics*, Oxford, 1935 (2nd edition).

[2] E. H. Kemble, *The Fundamental Principles of Quantum Mechanics*, New York and London, 1937.

[3] B. v. Sz. Nagy, *Spektraldarstellung linearer Transformationen des Hilbertschen Raumes*, Ann Arbor, 1947.

[4] J. v. Neumann, "Allgemeine Eigenwerttheorie Hermitescher Funktionaloperatoren," *Mathematische Annalen*, vol. 102 (1930), pp. 49-131.

[5] ———, *Mathematische Grundlagen der Quantenmechanik*, Berlin, 1932.

[6] C. R. Putnam, "On commutators of bounded matrices," *American Journal of Mathematics*, vol. 73 (1951), pp. 127-131.

[7] ———, "On normal operators in Hilbert space," *ibid.*, vol. 73 (1951), pp. 357-362.

[8] F. Rellich, "Der Eindeutigkeitssatz für die Lösungen der quantenmechanische Vertauschungsrelationen," *Göttinger Nachrichten*, 1946, pp. 107-115.

[9] M. H. Stone, *Linear Transformations in Hilbert Space*, New York, 1932.

[10] H. Wielandt, "Ueber die Unbeschränkheit der Operatoren der Quantenmechanik," *Mathematische Annalen*, vol. 121 (1949), p. 21.

[11] A. Wintner, *Spektraltheorie der unendlichen Matizen*, Leipzig, 1929.

[12] ———, "The unboundedness of quantum-mechanical matrices," *Physical Review*, vol. 71 (1947), pp. 738-739.

[13] ———, "On Dirac's theory of continuous spectra," *ibid.*, vol. 73 (1948), pp. 781-785.

ON NON-OSCILLATORY LINEAR DIFFERENTIAL EQUATIONS OF SECOND ORDER.*

By PHILIP HARTMAN.**

1. *Necessary conditions.* In the differential equation

$$y'' + q(t)y = 0, \tag{1}$$

let $q = q(t)$ be real-valued and continuous on $0 \leq t < \infty$. Only real solutions $y = y(t) \not\equiv 0$ will be considered below. The equation (1) will be called non-oscillatory if every solution has only a finite number of zeros on $0 \leq t < \infty$.

(I) *If* $q(t)$ *satisfies*

$$q(t) \to 0, \text{ as } t \to \infty, \tag{2}$$

or, more generally, if

$$\underset{0<v<\infty}{\text{l. u. b.}} \left| \int_u^{u+v} q(t)\,dt \right| / (1+v) \to 0, \text{ as } u \to \infty, \tag{3}$$

then a necessary condition that (1) *be non-oscillatory is that either*

$$\int_0^T q(t)\,dt \to -\infty, \text{ as } T \to \infty, \tag{4}$$

or

$$\int_0^\infty q(t)\,dt = \lim_{T\to\infty} \int_0^T q(t)\,dt \text{ exists} \tag{5}$$

(*as a finite limit*).

In the simple case where the assumption (2) or (3) is replaced by $q \geq 0$, it is known that (5) is a necessary condition (Wintner [9], p. 97). In fact, without any restriction on $q(t)$, it is known that

$$\liminf_{T\to\infty} T^{-1} \int_0^T \left(\int_0^t q(s)\,ds \right) dt < \infty$$

* Received May 21, 1951.

** John Simon Guggenheim Memorial Foundation Fellow, on leave of absence from The Johns Hopkins University.

is a necessary condition (Wintner [10]). The following theorem contains this result (but not that of (I)):

(II) *A necessary condition that* (1) *be non-oscillatory is that either*

$$\liminf_{T\to\infty} T^{-1} \int_0^T \left(\int_0^t q(s)\,ds \right) dt = -\infty \tag{6}$$

or

$$\lim_{T\to\infty} T^{-1} \int_0^T \left(\int_0^t q(s)\,ds \right) dt = M \text{ exists} \tag{7}$$

(as a finite limit).

The proof of (II) will show that, in the case (7), the mean value M exists "absolutely" and, in fact,

$$\lim_{T\to\infty} T^{-1} \int_0^T |\, Q(t)\,|^2 dt = 0, \tag{8}$$

where

$$Q(t) = M - \int_0^t q(s)\,ds. \tag{9}$$

Using the methods of [3], where however it is assumed that (5) holds and (9) reduces to $Q(t) = \int_t^\infty q(s)\,ds$, it can be shown that in the case (7) the function (9) must satisfy

$$\int_0^\infty \exp\left(-A \int_0^t Q(s)\,ds\right) dt = \infty \tag{10}$$

for $0 < A \leq 4$; but, according to Wintner ([11], § 7), not necessarily for $A > 4$. (That (10) must hold for $A = 2$ when (5) is assumed was first proved by Wintner [11], § 7.)

(II bis) *The case* (6) *of* (II) *is compatible with each of the possibilities*:

$$\limsup_{T\to\infty} T^{-1} \int_0^T \left(\int_0^t q(s)\,ds \right) dt \text{ is } -\infty, \text{ finite or } +\infty. \tag{11}$$

Of course, (I) rules out the last two possibilities when (2) or (3) holds. These possibilities are also ruled out by somewhat milder assumptions.

(III) *If $q(t)$ is half-bounded on $0 \leq t < \infty$, or more generally, if there exists a positive constant ϵ such that*

$$(12) \qquad \int_u^{u+v} q(t)\,dt \text{ is half-bounded for } 0 \leq u < \infty, 0 \leq v \leq \epsilon,$$

then a necessary condition that (1) *be non-oscillatory is that either* (4) *or* (7) *hold.*

2. *Auxiliary theorems.* The proof of (II) will depend on the following:

(A) *If* (1) *is non-oscillatory, then a necessary and sufficient condition that*

$$(13) \qquad \int^{\infty} (y'/y)^2 dt < \infty$$

hold for one (and/or every) solution $y = y(t)$ of (1) *is that* (7) *hold.*

This implies the result of Wintner ([11], § 6) that a sufficient condition that (13) hold for every solution of (1) is that (5) be valid. In view of (II), it is seen that even

$$(14) \qquad \liminf_{T\to\infty} T^{-1} \int_0^T \left(\int_0^t q(s)\,ds \right) dt > -\infty$$

is sufficient (since (14) implies (7), when (1) is non-oscillatory).

The proof of (I) will depend on the following:

(B) *If* (1) *is non-oscillatory, then* (3) *is necessary and sufficient in order that*

$$(15) \qquad y'(t)/y(t) \to 0, \text{ as } t \to \infty,$$

hold for one (and/or every) solution $y = y(t)$ of (1).

This implies the result of Wintner ([11], § 6) that a sufficient condition that (15) hold for every solution of (1) is that (5) be valid.

The theorem (B) is an analogue of a theorem in [2], § 11, where $q(t)$ in (1) is replaced by $q(t) - 1$, and (15) by $y'/y \to \pm 1$, as $t \to \infty$ (but where it is asserted, not assumed, in one half of the theorem that (1) is non-oscillatory.)

Theorem (B) implies the result of Perron [6] that (15) holds for every solution of (1) if (2) holds and $q(t) \leq 0$. Actually, Perron treats the more general equation

$$y'' + p(t)y' + q(t)y = 0, \qquad (p, q \text{ real}) \tag{16}$$

and his assumptions are that (2) holds, $q(t) \leq 0$,

$$p(t) \to 0, \text{ as } t \to \infty, \tag{17}$$

and $p(t) \leq 0$. Since there is no loss of generality in assuming that $p(t)$ is locally smooth, it follows from (B) and the variation of constants

$$y = z \exp(-\tfrac{1}{2} \int_0^t p(s)\,ds),$$

which gives the differential equation $z'' + (q - p^2/4 - p'/2)z = 0$ and the relation $y'/y = z'/z - p/2$, that if (17) is assumed, then (B) remains true if "(16)" is read in place of "(1)".

Since (16) is non-oscillatory when $q \leq 0$, it follows that Perron's assumption $p \leq 0$ is superfluous. This fact has already been noted by Hamel [1] who, under the assumptions (2) and (17), gave necessary and sufficient conditions that (16) be non-oscillatory and that its solutions satisfy (15). But it is clear from the extension of (B) to the differential equation (16) that Hamel's conditions must then be merely necessary and sufficient conditions that (16) be non-oscillatory. From this point of view, Hamel's criterion, based on a variational principle, that (16) be non-oscillatory has nothing to do with his assumptions (2) and (17). This is clear from a glance at the first part of his proof. His criterion has been generalized by Putnam [7].

3. *Proof of* (B). If $y = y(t)$ is any solution of (1), then $y(t)$ does not vanish for large t and its logarithmic derivative

$$x = y'/y \tag{18}$$

satisfies the Riccati differential equation

$$x' + x^2 + q(t) = 0 \tag{19}$$

(for those t for which (18) is defined). Hence, (B) is a consequence of the case $r(x) = -x^2$ of the following lemma, which is an analogue of the theorem (II*) in [2], § 9.

LEMMA 1. *Let $r(x)$ be defined and continuous for $-\infty < x < \infty$ and have the properties that*

$$r(0) = 0, \tag{20}$$

and that there exists, on $0 \leq x < \infty$, *a monotone (non-decreasing) function* $R(x)$ *satisfying*

(21) $$|r(x)| \geq R(|x|) > 0 \text{ for } 0 < |x| < \infty$$

and

(22) $$\int^{\infty} dx/R(x) < \infty.$$

Let $x = x(t)$ *be defined for large* t *and be a solution of*

(23) $$x' = r(x) - q(t).$$

Then (3) *is necessary and sufficient in order that*

(24) $$x(t) \to 0, \text{ as } t \to \infty.$$

Proof. The necessity of (3) follows from the inequality

$$\left| \int_u^{u+v} q(t)\,dt \right| \leq |x(u+v) - x(u)| + \int_u^{u+v} |r(x(t))|\,dt,$$

where (20), (21) and (24) imply that $r(x(t)) \to 0$, as $t \to \infty$.

In order to prove the sufficiency of (3), it will first be shown that (3) implies

(25) $$\liminf_{t\to\infty} x(t) \leq 0 \leq \limsup_{t\to\infty} x(t).$$

Suppose, if possible, that the second inequality does not hold; so that

(26) $$x(t) \leq -c < 0$$

for large t, say for $a \leq t < \infty$. Then, by (21) and (23), for $a \leq u < +v < \infty$,

(27) $$|x(u+v) - x(u)| \geq R(c)v - \left| \int_u^{u+v} q(t)\,dt \right|.$$

Clearly, (3) implies that, if u is sufficiently large and fixed, then the right side of (27) tends to ∞, as $v \to \infty$. Hence (26) shows that $x(t) \to -\infty$, as $t \to \infty$. From (23),

$$x(t) = \int_a^t r(x(s))\,ds - \int_a^t q(s)\,ds + x(a).$$

Since (21) and (22) imply that $|r(x(t))| \to \infty$, as $t \to \infty$, the condition (3) and the last formula line imply $r(x(t)) \to -\infty$, as $t \to \infty$. Hence (3) gives

9

$$-x(t) \geq \tfrac{1}{2} \int_a^t R(-x(s))\,ds \text{ for large } t.$$

Since R is monotone, it follows that

$$2S' \geq R(S), \text{ where } S(t) = \tfrac{1}{2} \int_a^t R(-x(s))\,ds$$

tend to ∞, as $t \to \infty$. A quadrature gives

$$(28) \qquad 2 \int^{S(t)} dS/R(S) \geq t + \text{const.}$$

If $t \to \infty$, this relation contradicts (22). This proves the second inequality in (25); the first is proved similarly.

Suppose, if possible, that the limit superior, as $t \to \infty$, of $x(t)$ is positive, say has the value $3c > 0$. Then, by (25), there exist large values of u and $u + v(> u)$ such that $c \leq x(t) \leq 2c$ for $u \leq t \leq u$ and that $x(u+v) - x(u) = \pm c$, where $\pm$ is $-\operatorname{sgn} r(c)$. Then an analogue of (27) gives

$$\left| \int_u^{u+v} q(t)\,dt \right| \geq c + R(c)v \geq c_0(1+v),$$

where $c_0 = \min(c, R(c)) > 0$. But this contradicts (3). Hence, the limit superior of $x(t)$ is 0; similarly, it is proved that the limit inferior is 0. This completes the proof of (24) and of Lemma 1.

4. *Proof of* (A). If $y = y(t)$ is any solution of (1), then its logarithmic derivative (18) satisfies the Riccati equation (19) for large t, say for $a \leq t < \infty$. An integration of (19) gives

$$(29) \qquad x(t) + \int_a^t x^2(s)\,ds = x(a) - \int_a^t q(s)\,ds$$

for $t \geq a$.

Suppose that (13) holds, it will be shown that (7) holds. The assumption (13) implies that (29) can be writen in the form

$$(30) \qquad x(t) - \int_t^\infty x^2(s)\,ds = M - \int_0^t q(s)\,ds,$$

where

$$(31) \qquad M = x(a) - \int_a^\infty x^2(s)\,dc + \int_0^a q(s)\,ds.$$

Since (13) implies

$$T^{-1}\int_a^t x^2(t)\,dt \to 0, \text{ as } T \to \infty,$$

and

$$T^{-1}\int_a^T (\int_t^\infty x^2(s)\,ds)^2 dt \to 0, \text{ as } T \to \infty,$$

the relation (8) follows from (9) and (30). In particular, (7) holds (Hölder), as was to be proved.

In order to prove the converse, note that (29) and a quadrature give

$$(32) \quad t^{-1}\int_a^t x(s)\,ds + t^{-1}\int_a^t (\int_a^s x^2(u)\,du)\,ds$$
$$= t^{-1}x(a)(t-a) - t^{-1}\int_a^t (\int_a^s q(u)\,du)\,ds.$$

The assumption (7), or even (14), means that the right side is bounded from above, as $t \to \infty$. If (13) fails to hold, then

$$-t^{-1}\int_a^t x(s)\,ds \geq \tfrac{1}{2}t^{-1}\int_a^t (\int_a^s x^2(u)\,du)\,ds \text{ for large } t.$$

Then Schwarz's inequality

$$| t^{-1}\int_a^t x(s)\,ds | \leq (t^{-1}\int_a^t x^2(s)\,ds)^{\frac{1}{2}}$$

shows that

$$4t\int_a^t x^2(u)\,du \geqq (\int_a^t (\int_a^s x^2(u)\,du)\,ds)^2 \text{ for large } t.$$

This is equivalent to

$$4tS' \geq S^2, \text{ where } S(t) = \int_a^t (\int_a^s x^2(u)\,du)\,ds$$

tends to ∞, as $t \to \infty$. A quadrature of this inequality gives

$$4 \int^{S(t)} ds/S^2 \geq \log t + \text{const.}$$

Since this is impossible for large t, it follows that (7), or even (14), implies (13) for every solution $y(t)$ of (1). This completes the proof of (A).

5. *Proof of* (II). Suppose that (1) is non-oscillatory and that (6) does not hold (that is, that (14) holds). It must be shown that the alternative, (7), holds. By the proof just completed, (14) implies (13) for every solution $y(t)$ of (1). But (13), for one solution, implies (7), by (A). This proves (II).

6. *Proof of* (III). If (4) fails to hold, then the integral in (4) exceeds some fixed constant when $t = t_n$, where $t_1 < t_2 < \cdots$ is some unbounded sequence of t-values. The half-boundedness of q, or the more general condition (12), implies the existence of constants C and $\epsilon > 0$ such that

$$\int_0^t q(s)\,ds > -C \tag{33}$$

for $t_n < t < t_n + \epsilon$ or $t_n - \epsilon < t < t_n$, where $n = 1, 2, \cdots$, and C, ϵ are independent of n. The assertion (III) is contained, therefore, in the following:

LEMMA 2. *If* (1) *is non-oscillatory and if there exists a constant C such that* (33) *holds for all t on a set E of infinite measure, then* (7) *is valid.*

Proof: By virtue of (A), it is sufficient to show that (13) holds for a solution $y = y(t)$ of (1). The Ricatti equation (19) and its integrated form (29) are satisfied for $x = y'/y$ for large t, say $a \leq t < \infty$. By (33), the right side of (29) is bounded from above if t is on the set E. Hence, if (13) fails to hold, a modification of the arguments between (27) and (28) leads to the contradiction

$$2 \int_{E(T)} dS(t)/S^2(t) \geq \text{measure of } E(T),$$

where T is large, $E(T)$ denotes the portion of E on $T \leq t < \infty$, and

$$S(t) = \tfrac{1}{2} \int_a^t x^2(s)\,ds \to \infty, \text{ as } t \to \infty.$$

This proves Lemma 2.

7. *Proof of* (I). If (1) is non-oscillatory and (2), or more generally (3), hold then, either (4) or (7) holds, by (III). It remains to show that when (7) holds, (5) is valid.

Suppose that (7) holds, then (13) is true for every solution of (1), by (A). In addition, (15) is true, by (B). The equation (30), which is valid for large t, then implies (5). This proves (I).

8. *Other necessary conditions.* Changes of the variables in (1), followed by applications of the above theorems lead to different necessary conditions that (1) be non-oscillatory. For example, if the new independent variable, $s = t^\gamma$, where $\gamma > 0$, and the new dependent variable $z = t^{\frac{1}{2}(\gamma-1)}y$ are introduced, then (1) becomes

$$d^2z/ds^2 + \gamma^{-2}[q(t)t^{2-2\gamma} - (1-\gamma^2)t^{-2\gamma}/4]z = 0, \text{ where } t = s^{1/\gamma}.$$

An application of (II) shows that if

$$\liminf_{T\to\infty} \int^T q(t)t^{1-\gamma}dt > -\infty,$$

then a necessary condition that (1) be non-oscillatory is that the mean value

$$(34) \qquad \lim_{T\to\infty} T^{-\gamma} \int^T t^{\gamma-1}\left(\int^t q(u)u^{1-\gamma}du\right)dt = M \text{ exist}$$

(absolutely). In particular, if $q \geqq 0$, then

$$\int^\infty q(t)t^{1-\gamma}dt < \infty \text{ for all } \gamma > 0.$$

This result can be deduced from Wintner [9], p. 97; [10]; [11]; Hille [4], p. 243; and Leighton [5]. If $\gamma > 0$ is fixed and $q(t)t^{2-2\gamma} \to 0$, as $t \to \infty$, then (I) show that $\lim \int^T q(t)t^{1-\gamma}dt$, as $T \to \infty$, exists either as a finite number or as $-\infty$. Finally, if $\gamma \geqq 1$ and $q(t)t^{1-\gamma}$ is half-bounded, as $t \to \infty$, then it can be shown, by (III), that either (34) holds or $\int^T q(t)t^{1-\gamma}dt \to -\infty$, when (1) is non-oscillatory.

9. *Proof of* (II bis). The first possibility in (11) is realized, for example, by $q(t) \equiv -1$.

In order to obtain examples illustrating the other two possibilities of (11), suitable continuously differentiable functions $x = x(t)$ will be constructed on $0 \leq t < \infty$. The function $q(t)$ will then be defined, in terms of $x(t)$, by the Riccati equation (19). It then follows that the corresponding differential equation (1) is non-oscillatory; in fact, $y(t) = \exp \int_0^t x(s)\,ds$ is a non-vanishing solution of (1) for $0 \leq t < \infty$. This type of construction of counter-examples in the theory of the differential equation (1) has been given in [8], p. 394 and [2], § 21.

The function $x = x(t)$ to be defined now will be smooth except for a sequence of discontinuities, which tend to ∞. It will be clear from the proof that these discontinuities can easily be avoided.

Since (32) follows from (19), where a is arbitrary, it must be shown that if $X(t)$ denotes the left side of (32), then for suitable choices of $x = x(t)$,

$$\limsup_{t\to\infty} X(t) = \infty \tag{35}$$

and

$$\liminf_{t\to\infty} X(t) \text{ is finite or } -\infty. \tag{36}$$

The relation (35) is equivalent to (6); while the first and second alternatives of (36) are equivalent to the second and third, respectively, of (11).

Let $\epsilon_1, \epsilon_2, \cdots$ be a sequence of positive numbers satisfying

$$\alpha > \epsilon_n \to 0, \text{ as } n \to \infty, \qquad (\alpha = 1/e). \tag{37}$$

Define $x(t)$, on $(a =)\,1 - \alpha \leq t < \infty$, by placing $x(t) = 1/(n-t)\log(n-t)$ or $x(t) = 0$ according as $n - \alpha \leq t < n - \epsilon_n$ or $n - \epsilon_n \leq t < n + 1 - \alpha$ for $n = 1, 2, \cdots$. Then

$$T^{-1}\int_a^T x(t)\,dt = -T^{-1}\sum_{k=1}^{n} \log\log \epsilon_k^{-1} \text{ for } n - \epsilon_n \leq T \leq n + 1 - \alpha. \tag{38}$$

The identity

$$\int_a^T \Big(\int_a^t x^2(s)\,ds\Big)\,dt = \int_a^T (T - s)\,x^2(s)\,ds$$

and a simple calculation show that, for $n - \epsilon_n \leq T \leq n + 1 - \alpha$,

$$(39) \qquad T^{-1} \int_a^T \left(\int_a^t x^2(s)\,ds \right) dt = \sum_{k=1}^{n} (1 - k/T) I(\epsilon_k) + T^{-1} \sum_{k=1}^{n} (1 + \log^{-1} \epsilon_k),$$

where

$$(40) \qquad I(\epsilon) = \int_\epsilon^a dt/t^2 \log^2 t \sim 1/\epsilon \log^2 \epsilon, \text{ as } \epsilon \to 0.$$

If $T = n + 1 - \alpha$, then $1 - k/T \geq (1 - \alpha)/T$ for $k = 1, \cdots, n$. Also, the second term on the right of (39) is positive, by (37). Hence, $X(T)$, which is the sum of (38) and (39), exceeds

$$T^{-1} \sum_{k=1}^{n} (1 - \alpha) I(\epsilon_k) - \log\log \epsilon_n^{-1}$$

for $T = n + 1 - \alpha$. In view of (40), this implies

$$\underset{(n-a,\, n+1-a)}{\text{l. u. b.}}\ X(t) \to \infty, \text{ as } n \to \infty;$$

and so, (35) holds.

Let $T = n$, then the contribution of the interval $n - \alpha \leq t \leq n - \epsilon_n$ to $X(T)$ is $n^{-1}(1 + \log^{-1} \epsilon_n - \log\log \epsilon_n^{-1})$. This expression tends to 0 or to $-\infty$ according as ϵ_n tends to 0 sufficiently slowly or sufficiently rapidly, as $n \to \infty$. The contribution of $a \leq t < n - \alpha$ to $X(T)$, where $n - \alpha \leq t \leq n + 1 - \alpha$, depends only on $\epsilon_1, \cdots, \epsilon_{n-1}$ (and not on ϵ_n) and exceeds

$$T^{-1} \sum_{k=1}^{n-1} [(1 - \alpha) I(\epsilon_k) - \log\log \epsilon_k^{-1}] \to \infty, \text{ as } n \to \infty.$$

Consequently, if $c_1, c_2, \cdots$ is any given sequence of numbers, which is bounded from above, then there exists a suitable sequence $\epsilon_1, \epsilon_2, \cdots$ satisfying (37) and

$$\underset{(n-a,\, n+1-a)}{\text{g. l. b.}}\ X(t) \sim c_n, \text{ as } n \to \infty;$$

in particular,

$$\liminf_{t \to \infty} X(t) = \liminf_{n \to \infty} c_n.$$

On choosing c_n independent of n or $c_n = -n$, the alternatives of (36) are realized. This completes the proof of (II bis).

PARIS, FRANCE.

REFERENCES.

[1] G. Hamel, "Ueber das infinitäre Verhalten der Integrale einer linearen Differentialgleichung zweiter Ordnung, wenn die characterische Gleichung zwei gleiche Wurzeln hat," *Mathematische Zeitschrift*, vol. 1 (1918), pp. 220-228.

[2] P. Hartman, "Unrestricted solution fields of almost-separable differential equations," *Transactions of the American Mathematical Society*, vol. 63 (1948), pp. 560-580.

[3] ———, "On linear second order differential equations with small coefficients," *American Journal of Mathematics*, vol. 73 (1951), pp. 955-962.

[4] E. Hille, "Non-oscillation theorems," *Transactions of the American Mathematical Society*, vol. 64 (1948), pp. 234-252.

[5] W. Leighton, "The detection of the oscillation of solutions of a second order linear differential equation," *Duke Mathematical Journal*, vol. 17 (1950), pp. 57-62.

[6] O. Perron, "Ueber das infinitäre Verhalten der Integrale einer linearen Differentialgleichung zweiter Ordnung, wenn die characteristische Gleichung zwei gleiche Wurzeln hat," *Sitzungsberichte der Heidelberger Akademie*, 1917, Abhandlung 9.

[7] C. R. Putnam, "An oscillation criterion involving a minimum principle," *Duke Mathematical Journal*, vol. 16 (1949), pp. 633-636.

[8] A. Wintner, "The adiabatic linear oscillator," *American Journal of Mathematics*, vol. 68 (1946), pp. 385-397.

[9] ———, "On the Laplace-Fourier transcendents occurring in mathematical physics," *ibid.*, vol. 69 (1947), pp. 87-97.

[10] ———, "A criterion of oscillatory stability," *Quarterly Journal of Applied Mathematics*, vol. 7 (1949), pp. 115-117.

[11] ———, "On the non-existence of conjugate points," *American Journal of Mathematics*, vol. 73 (1951), pp. 368-380.

ON THE DIRICHLET SERIES ASSOCIATED WITH RAMANUJAN'S τ-FUNCTION.*

By George Shapiro.[1]

I. Introduction. The Ramanujan function $\tau(n)$ is defined by the identity

$$x \prod_{n=1}^{\infty} (1-x^n)^{24} = \sum_{n=1}^{\infty} \tau(n) x^n;$$

$t(n)$, by the relation

$$t(n) = \tau(n) n^{-11/2}.$$

It is known [7], that

$$t(n) = O(n^{3/10}), \tag{1.1}$$

while Ramanujan has conjectured that

$$t(n) = O(n^{\epsilon}) \text{ for every } \epsilon > 0. \tag{1.2}$$

The function $t(n)$ is multiplicative; i. e.,

$$t(mn) = t(m)t(n) \text{ if } (m, n) = 1. \tag{1.3}$$

Also,

$$t(p^{\lambda}) = t(p)t(p^{\lambda-1}) - t(p^{\lambda-2}), \qquad \lambda = 2, 3, \cdots. \tag{1.4}$$

Put

$$Z(s) = \sum_{n=1}^{\infty} t(n) n^{-s}. \tag{1.5}$$

This series is known to converge for $\sigma > 2/5$ and to have as abscissa of absolute convergence $\sigma = 1$. Therefore, it follows from (1.3), (1.4) and (1.5) that the Euler representation of $Z(s)$ takes the special form

$$Z(s) = \prod_{p} (1 - t(p)p^{-s} + p^{-2s})^{-1} \text{ for } \sigma > 1, \tag{1.6}$$

and that the product is absolutely convergent there.

Wilton [9] has shown that

$$(2\pi)^{-s}\Gamma(s+11/2)Z(s) = (2\pi)^{s-1}\Gamma(13/2-s)Z(1-s), \tag{1.7}$$

* Received June 5, 1950; revised January 9, 1952.

[1] The author was a Predoctoral Fellow of the Atomic Energy Commission during the research leading to this paper.

and that $Z(s)$ has infinitely many zeros on the line $\sigma = 1/2$. From (1.6) it follows that $Z(s)$ has no zeros in the half plane $\sigma > 1$; the functional equation, (1.7), then shows that $Z(s)$ has "trivial" zeros at $s = -11/2$, $-13/2, \cdots$ and no other zero with negative real part. Moreover, Rankin [6] has shown that there are no zeros on the lines $\sigma = 0$ and $\sigma = 1$. The "Riemann" hypothesis for $Z(s)$ is that all non-trivial zeros of $Z(s)$ lie on the line $\sigma = 1/2$.

Let G be the (simply connected) region obtained from the half plane $\sigma > 1/2$ by deleting all horizontal segments

$$\sigma + it_0, \qquad 1/2 < \sigma \leq \sigma_0,$$

where $\sigma_0 + it_0$ is a zero, if any exist, of $Z(s)$ in $\sigma > 1/2$. By $\operatorname{Log} Z(s)$ in $\sigma > 1/2$ is meant the analytic continuation into G of $\operatorname{Log} Z(s)$ in $\sigma > 1$, where $\operatorname{Log} Z$ is that branch of the logarithm which is real for real positive Z.

Put

$$Z_n(s) = \prod_{k=1}^{n} (1 - t(p_k) p_k^{-s} + p_k^{-2s})^{-1} \text{ for } \sigma > 1/2.$$

In II it will be shown by a method similar to that used by Bohr in [1], that for every fixed $\sigma > 13/20$, the sequences $\{Z_n(s)\}$ and $\{\operatorname{Log} Z_n(s)\}$ converge in relative measure to $Z(s)$ and $\operatorname{Log} Z(s)$, respectively. In III, the asymptotic distribution functions of Z and of $\operatorname{Log} Z$ will be investigated. (See [5] for definitions of the concepts used in III; much of III is analogous to the corresponding investigation of $\zeta(s)$ in [5].) It will be shown that both asymptotic distribution functions are absolutely continuous with densities of class C^{∞}, and that for the former, the density vanishes only at $x = 0$ when $13/20 < \sigma < 1$. This is an indication towards the Riemann hypothesis. On the basis of Ramanujan's conjecture (1.2), the same results can be obtained for $\sigma > 1/2$.

II. Convergence in relative measure. In the first part of II, we assume the Ramanujan conjecture. Let $\nu(\sigma) = \text{g.l.b. } \lambda$ for which

$$\int_{-T}^{T} |Z(\sigma + it)|^2 dt = O(T^{\lambda}), \qquad (T \to \infty).$$

Carlson [3] has shown that $\nu(\sigma)$ is a convex function, and that if

$$q = \text{g.l.b. } \{1 - (1 - v)/\nu(v)\}, \tag{2.1}$$

then

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |Z(\sigma + it)|^2 dt$$

exists and is finite for $\sigma > q$. From (1.7) and the asymptotic formula for the Γ-function, it follows that

$$\int_{-T}^{T} |Z(s)|^2 dt = O(T^{2-4\sigma} \int_{-T}^{T} |Z(1-s)|^2 dt) = O(T^{3-4\sigma}),$$

if

$$\int_{-T}^{T} |Z(1-\sigma-it)|^2 dt = O(T). \tag{2.2}$$

When $\sigma < 0$, the Dirichlet series for $Z(1-s)$ is absolutely convergent, so that (2.2) holds. Therefore,

$$\nu(\sigma) \leq 3 - 4\sigma \text{ for } \sigma \leq 0,$$

and in particular $\nu(0) \leq 3$, so that $q \leq 2/3$. Therefore, (2.2) is valid for $\sigma < 1/3$, so that

$$\nu(\sigma) \leq 3 - 4\sigma \text{ for } \sigma \leq 1/3,$$

and $\nu(1/3) \leq 5/3$, which implies that $q \leq 3/5$. On repeating this procedure, one obtains $q \leq (n+1)/(2n+1)$. Letting $n \to \infty$, we have $q \leq 1/2$. But $q < 1/2$ implies, by a theorem of Carlson [2], that

$$\sum_{1}^{\infty} t^2(n) n^{-2\sigma} < \infty, \quad q < \sigma < 1/2. \tag{2.3}$$

On the other hand,

$$\sum_{1}^{n} t^2(m) \sim an, \qquad a \neq 0$$

(Rankin [7]), so that the series in (2.3) diverges. Hence $q = 1/2$.

If (1.2) is not assumed, then in (2.1), the number 1 must be replaced by 23/20; the corresponding result is that $q \leq 13/20$. Put $\sigma^* = 1/2$ or 13/20 according as (1.2) is or is not true.

By virtue of the definition of $Z_n(s)$,

$$|Z_n(\sigma+it)|^{-1} \leq \prod_{k=1}^{n} (1 + |t(p_k)| p_k^{-\sigma} + p_k^{-2\sigma}) = O(1),$$

as $|t| \to \infty$. Therefore,

$$\int_{-T}^{T} |Z(s)/Z_n(s)|^2 dt = O(T) \text{ as } T \to \infty, \text{ when } \sigma > \sigma^*.$$

From the inequality $(a-1)^2 \leq 2a^2 + 2$, it follows that

$$\int_{-T}^{T} |Z(s)/Z_n(s) - 1|^2 dt = O(T), \text{ when } \sigma > \sigma^*;$$

so that by the theorem of Carlson in [3] referred to above,

$$\lim_{T\to\infty}\int_{-T}^{T}| Z(s)/Z_n(s) -1 |^2dt = \sum_{m=1}^{\infty} c^2_n(m)m^{-2\sigma} \text{ for } \sigma > \sigma^*,$$

where $\sum_{m=1}^{\infty} c_n(m)m^{-s} = Z(s)/Z_n(s) - 1$ for $\sigma > 1$, say. But for $\sigma > 1$,

$$Z(s)/Z_n(s) = \prod_{k=n+1}^{\infty}(1 - t(p_k)p_k^{-s} + p_k^{-2s})^{-1},$$

so that

$$c_n(1) = c_n(2) = \cdots = c_n(n) = 0,$$

while $| c_n(m)| \leq | t(m)|$ for all m. Thus, for any fixed $\sigma > \sigma^*$,

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}| Z(s)/Z_n(s) -1 |^2dt \leq \sum_{n+1}^{\infty} t^2(m)m^{-2\sigma} \to 0 \text{ as } n\to\infty.$$

Therefore, if $\text{meas}_T\{| Z(s)/Z_n(s) - 1 | > \epsilon\}$ denotes the Lebesgue measure of the set of t- vaules for which $-T \leq t \leq T$ and $Z(s)/Z_n(s) - 1 | > \epsilon$, then

$$\lim_{n\to\infty}\limsup_{T\to\infty} \text{meas}_T\{|Z(s)/Z_n(s) - 1 | > \epsilon\}/2T = 0.$$

Since $| \text{Log}\, Z - \text{Log}\, Z_n | = | \text{Log}\,[1 + (Z/Z_n - 1)]| < | Z/Z_n - 1 |$, when $| Z/Z_n - 1 |$ is sufficiently small, it follows that $\{\text{Log}\, Z_n(s)\}$ converges in relative measure to $\text{Log}\, Z(s)$ for every fixed $\sigma > \sigma^*$; furthermore, since $Z_n(s) = O(1)$ as $| t | \to \infty$, it follows as before, that

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}| Z(s) - Z_n(s) |^2dt \leq \sum_{n+1}^{\infty} 4t^2(m)m^{-2\sigma}, \qquad \sigma > \sigma^*;$$

so that for every fixed $\sigma > \sigma^*$, $\{Z_n(s)\}$ converges in relative measure to $Z(s)$.

III. The distribution functions. We shall prove

THEOREM 3. 1. *Put*

$$F_k(z) = -\text{Log}(1 - t(p_k)z + z^2);$$

for fixed σ, let S_k denote the curve

$$x_k(\theta) = \xi_k(\theta) + i\eta_k(\theta) = F_k(p_k^{-\sigma}e^{2\pi i\theta}),$$

and let ϕ_k be the distribution function determined by this parametric representation of S_k; that is, of E is a Borel set in the x-plane, then $\phi_k(E)$ is the θ measure of ES_k. Let $\Lambda(y, \phi_k)$ denote the Fourier transform of ϕ_k:

$$\Lambda(y,\phi_k) = \int_0^1 \exp(ig_{\tau,k}(\theta)| y |)\,d\theta,$$

where $y = (|y| \cos 2\pi\tau, |y| \sin 2\pi\tau)$ *and* $g_{\tau,k} = \xi_k(\theta) \cos 2\pi\tau + \eta_k(\theta) \sin 2\pi\tau$. *Then, for* $\sigma > \sigma^*$,

$$\Lambda(y, \phi_k) = O(|y|^{-\frac{1}{2}}) \text{ as } |y| \to \infty,$$

for infinitely many k.

Proof. For convenience we drop the subscript k. We apply Theorem 12 of [5]. It is clear that ξ and η possess continuous second derivatives; we must show that S is convex and that $g_\tau''(\theta)$ has, for every fixed τ, exactly two zeros θ on S.

An application of the Study Rundungsschranke ([8], p. 109) gives the result that a sufficient condition that S be convex is that

(3.1) $\qquad 13p^{-\sigma} \leq |t(p)| \leq p^{\sigma}/2$ and p is sufficiently large.

In fact, the Study theorem states that S is convex if $F'(0) \neq 0$ and

$$\Re[1 + zF''/F'] > 0 \text{ for } |z| = p^{-\sigma}.$$

Now $F'(0) = t(p)$, while

$$\begin{aligned} |(t(p) - 2z)(1 - t(p)z + z^2)|^2 \Re[1 + zF''/F'] = \\ (t^2(p) + 3t^2(p)p^{-4\sigma} + 8p^{-2\sigma} - 8p^{-4\sigma}) \\ - (t^3(p) + 6t(p) + t^3(p)p^{-2\sigma} + 8t(p)p^{-2\sigma} + 2t(p)p^{-4\sigma})p^{-\sigma} \cos 2\pi\theta \\ + (8t^2(p) + 16p^{-2\sigma})p^{-2\sigma} \cos^2 2\pi\theta - 8t(p)p^{-3\sigma} \cos^3 2\pi\theta, \end{aligned}$$

where $z = p^{-\sigma}e^{2\pi i\theta}$. Clearly it is sufficient to consider $t(p) > 0$ and $\cos 2\pi\theta > 0$. The last two terms of the preceding expression can be written in the form

$$8p^{-2\sigma}(t^2(p) + 2p^{-2\sigma} - t(p)p^{-\sigma} \cos 2\pi\theta) \cos^2 2\pi\theta.$$

This is non-negative, since $2|t(p)| p^{-\sigma} \leq t^2(p) + p^{-2\sigma}$. The last two members, $8p^{-2\sigma} - 8p^{-4\sigma}$, of the term independent of θ dominate the last two members of the coefficient of $\cos 2\pi\theta$,

$$8p^{-2\sigma} - 8p^{-4\sigma} \geq 8t(p)p^{-3\sigma} + 2t(p)p^{-5\sigma} \text{ if } p \geq 3,$$

by virtue of the right hand inequality of (3.1). Similarly, the first member, $t^2(p)$, of the term independent of θ dominates the first three members of the coefficient of $\cos 2\pi\theta$, for

$$t^2(p) - (t^3(p)p^{-\sigma} + 6t(p)p^{-\sigma} + t^3(p)p^{-3\sigma}) \geq t^2(p)/2 - 25t(p)p^{-\sigma}/4 \geq 0$$

follows from (3.1). Thus (3.1) is sufficient to assure the convexity of S; by a more tedious computation it may be shown that if $0 < |t(p)| \leq p^{\sigma}$,

and if $2p^{-\sigma} \leq |t(p)| \leq 4p^{-\sigma}$ *does not hold,* then S is convex. The right hand inequality of (3.1) is superfluous in view of (1.1) and $\sigma > \sigma^* \geq 1/2$; also (3.1) may be replaced by the condition

$$(3.2) \qquad |t(p)| \geq p^{-\frac{1}{4}} \text{ and } p \text{ sufficiently large.}$$

That some such condition is necessary is shown by the fact that if $t(p) = 3p^{-\sigma}$ and $z = p^{-\sigma}$, then $\Re[1 + zF''/F'] < 0$.

Infinitely many primes satisfy (3.2), for if not, the series (1.5) would converge absolutely for $\sigma > 1/2$. It remains to show that for those primes satisfying (3.2), $g_\tau''(\theta)$ has just two zeros. Put $\nu = \nu_k = -1$ if $t(p) < -2$, and $\nu = 1$ otherwise. If $\gamma = \gamma_k$ is the real or purely imaginary angle satisfying

$$2 \cos 2\pi\gamma = \nu t(p) \text{ and } 0 \leq \gamma < 1/2 \text{ or } i\gamma < 0,$$

then

$$(3.3) \qquad F(z) = 2 \sum_1^\infty \nu^n z^n \cos 2\pi n\gamma/n, \qquad |z| \leq p^{-\frac{1}{2}}.$$

Furthermore, it is easy to see that $F(p^{-\sigma}e^{2\pi i\theta}) = -\operatorname{Log}(1-\zeta_1)(1-\zeta_2)$, where $\zeta_1 = \nu p^{-\sigma}e^{2\pi i(\theta-\gamma)}$, $\zeta_2 = \nu p^{-\sigma}e^{2\pi i(\theta-\gamma)}$; so that

$$g_\tau''(\theta) = -4\pi^2\Re\{[\zeta_1(1-\zeta_1)^{-2} + \zeta_2(1-\zeta_2)^{-2}]e^{-2\pi i\tau}.$$

If $p \to \infty$ through the primes satisfying (3.2), then

$$\zeta_1, \zeta_2, (\zeta_2 e^{2\pi i\gamma} + \zeta_1 e^{-2\pi i\gamma})/\cos 2\pi\gamma, \quad (\zeta_2^2 e^{2\pi i\gamma} - \zeta_1^2 e^{-2\pi i\gamma})/\cos 2\pi\gamma$$

tend to 0 uniformly for all θ, by virtue of (1.1) and $\sigma > \sigma^*$. Therefore,

$$g_\tau''(\theta) = -8\pi^2\nu p^{-\sigma} \cos 2\pi\gamma \cos 2\pi(\theta - \tau) + o(1)$$

uniformly for all θ and τ, as $p \to \infty$ through those primes satisfying (3.2). Hence, the zeros of $[g_\tau''(\theta)/\nu p^{-\sigma} \cos 2\pi\gamma]$ cluster only at $\theta = \tau \pm 1/4$, the zeros of $\cos 2\pi(\theta - \tau)$. (It is understood that p is restricted to those primes satisfying (3.2).) In the same fashion it may be shown that the zeros of g''' are near $\theta = \tau$, $\tau + 1/2$. Since there is a zero of g''' between every two zeros of g'', it follows that if p is a sufficiently large prime satisfying (3.2), there is at most one zero of g'' near each of $\theta = \tau \pm 1/4$; inasmuch as a continuous derivative of a continuous periodic function has at least two zeros, it is clear that for all primes satisfying (3.2), $g_\tau''(\theta)$ has for every fixed τ precisely two zeros θ on S, which completes the proof of Theorem 3.1.

Next we show that for any fixed $\sigma > \sigma^*$, the infinite convolution $\phi_1 * \phi_2 \cdots$ is absolutely convergent. Now $x_k(\theta) \to 0$ uniformly for all θ, as $k \to \infty$; hence, S_k, the spectrum of ϕ_k, is contained in a fixed circle. If $c(\phi_k)$ and

$M(\phi_k)$ are the first and second moments of ϕ_k, respectively, then $c(\phi_k)=0$, so that the convergence of

$$\sum_1^\infty M(\phi_k)=4\sum_1^\infty\sum_1^\infty p_k^{-2m\sigma}\cos^2 2\pi m\gamma_k/m^2$$

is necessary and sufficient for the convergence and absolute convergence of the infinite convolution $\phi=\phi_1*\phi_2*\cdots$, by Theorems 5 and 6 of [5]. The convergence of the double series is obvious if Ramanujan's hypothesis is true. Otherwise, $\cos^2 2\pi m\gamma_k=O(p_k^{3m/5})$ as $m\to\infty$, uniformly in k. Hence

$$\sum_{k=1}^\infty\sum_{m=2}^\infty p_k^{-2m\sigma}\cos^2 2\pi m\gamma_k/m^2=O(1)\sum_{k=1}^\infty\sum_{m=2}^\infty p_k^{m(3/5-2\sigma)}/m^2,$$

the last series being convergent for $\sigma>11/20$; also

$$\sum_1^\infty p_k^{-2\sigma}\cos^2 2\pi\gamma_k=O(1)\sum t^2(p)p^{-2\sigma},$$

which converges for $\sigma>13/20$. Thus, the infinite convolution is absolutely convergent for $\sigma>\sigma^*$.

Since $|\Lambda(y,\phi_k)|\leq 1$ for all k, and $\Lambda(y,\phi_k)=O(|y|^{-1/2})$ for infinitely many k, we have $\Lambda(y,\phi)=\Lambda(y,\phi_1)\Lambda(y,\phi_2)\cdots=O(|y|^{-m})$ as $|y|\to\infty$, for every fixed m. Therefore (cf. [5], § 3), ϕ is absolutely continuous, and its density $D(x)$ has continuous partial derivatives of all orders.

There is an essential difference in the spectrum of ϕ if $\sigma>1$ and if $1>\sigma>\sigma^*$. In the former case, the spectrum is a bounded set. This follows from Theorem 3 of [5]. On the other hand, it will now be shown that if $1>\sigma>\sigma^*$, the spectrum of ϕ is the entire plane. First $\sum|t(p)|p^{-\sigma}=\infty$, for if this were false it would follow from the product representation (1.6) of $Z(s)$ that $\sum|t(n)|n^{-\sigma}<\infty$, which is false. The following calculation of the spectrum for $1>\sigma>\sigma^*$ is modelled on the proof of the second part of Theorem 14 of [5]. Let $1>\sigma>\sigma^*$ be fixed. If C is a sufficiently large constant, then

$$|F_k(z)-t(p_k)z|\leq C\max(4,t^2(p_k))|z|^2 \text{ for } |z|=p_k^{-\sigma}.$$

This follows from the representation (3.3) of $F_k(z)$, since

$$|\cos 2\pi m\gamma|\leq|e^{-2\pi im\gamma}|$$

implies that

$$|F_k(z)-t(p_k)z|\leq 2|e^{-2\pi i\gamma}z|^2/(1-|e^{-2\pi i\gamma}z|).$$

Clearly, $\sum\max(4,t^2(p_k))p_k^{-2\sigma}<\infty$, since both $\sum p_k^{-2\sigma}$ and $\sum t^2(p_k)p_k^{-2\sigma}$ converge. Let $\epsilon>0$ be arbitrary, and fix m so large that

$$C \sum_{m+1}^{\infty} \max(4, t^2(p_k)) p_k^{-2\sigma} < \epsilon.$$

Let x_0 be any point of the x-plane, and let x_1 be an arbitrary point of the form

$$F_1(p_1^{-\sigma} e^{2\pi i\theta_1}) + \cdots + F_m(p_m^{-\sigma} e^{2\pi i\theta_m}).$$

Then, if $n > m$ is large enough, $\sum_{m+1}^{n} | t(p_k) | p_k^{-\sigma} > | x_0 - x_1 |$, and none of the numbers $| t(p_k) | p_k^{-\sigma}$, $k = m+1, m+2, \cdots, n$ is greater than the sum of the others. Therefore, we may choose $\theta_{m+1}, \theta_{m+2}, \cdots, \theta_n$ so that $\sum_{m+1}^{n} t(p_k) p_k^{-\sigma} e^{2\pi i\theta_k} = x_0 - x_1$, which implies that $| \sum_{1}^{n} F_k(p_k^{-\sigma} e^{2\pi i\theta_k}) - x_0 | < \epsilon$, so that the spectrum of ϕ is the entire plane.

It is true, moverover, that for any σ satisfying $1 > \sigma > \sigma^*$, the density of ϕ is positive for all x. For a proof of this, cf. [5], p. 60.

Remark. The same will be true for $\sigma = 1$ if $\sum | t(n) | / n = \infty$. If the latter series converges, the spectrum of ϕ for $\sigma = 1$ is a bounded set. We may sum up the preceding remarks in the following theorem.

THEOREM 3. 2. *For any fixed $\sigma > \sigma^*$, the infinite convolution $\phi_1 * \phi_2 * \cdots$ is absolutely convergent, $\phi = \phi_\sigma = \phi_1 * \phi_2 * \cdots$ is absolutely continuous, and its density $D_\sigma(x)$ possesses continuous partial derivatives of all orders. If $1 > \sigma > \sigma^*$, then $D_\sigma(x) > 0$ for all x. The spectrum is symmetric with respect to the real axis; the spectrum of ϕ_σ is a bounded domain if $\sigma > 1$, and is the whole x-plane if $1 > \sigma > \sigma^*$. For any fixed $\lambda > 0$, $D_\sigma(x) = O(e^{-\lambda |x|^2})$ as $| x | \to \infty$; this appraisal is valid for each partial derivative of $D_\sigma(x)$.*

The syymetry of the spectrum is obvious; a proof of the last assertion may be obtained by modifying the proof of Theorem 16 of [5] in the same way that the proof of part of Theorem 14 was modified to verify that for $1 > \sigma > \sigma^*$, the spectrum of ϕ is the whole plane.

Still following [5] we define $\bar{\phi}_\sigma$ by $\bar{\phi}_\sigma(E) = \phi_\sigma(\log E)$; the following may readily be obtained:

THEOREM 3. 3. *The distribution $\bar{\phi}_\sigma$ is symmetric with respect to the real axis; if $1 > \sigma > \sigma^*$, its spectrum is the entire plane; if $\sigma > 1$, the spectrum is a bounded set not containing the origin. This distribution is absolutely continuous with a density $\bar{D}_\sigma(x)$ which possesses continuous partial derivatives of all orders. For any fixed $\lambda > 0$, $\bar{D}_\sigma(x) = O(e^{-\lambda \log^2 |x|})$ as $| x | \to \infty$ or $| x | \to 0$; these appraisals are valid also for each partial derivative of $\bar{D}_\sigma(x)$. If $1 > \sigma > \sigma^*$, then $\bar{D}_\sigma(x) = 0$ if and only if $x = 0$.*

Finally, we have the analogues of Theorems 30 and 31 of [5]:

THEOREM 3.4. *The functions* $\operatorname{Log} Z(\sigma+it)$ *and* $Z(\sigma+it)$ *possess asymptotic distribution functions in the unrestricted case if* $\sigma>1$, *and in the restricted case if* $1\geqq\sigma>\sigma^*$. *These are, respectively, the distribution functions* ϕ_σ *and* $\bar{\phi}_\sigma$ *discussed above. The spectrum of* ϕ_σ *is the closure of the range of* $\operatorname{Log} Z(\sigma+it)$, *and the spectrum of* ϕ_σ *is the closure of the range of* $Z(\sigma+it)$ *for* $-\infty<t<\infty$.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] H. Bohr, "Zur Theorie der Riemann'schen Zetafunktion im kritischen Streifen," *Acta Mathematica*, vol. 40 (1916), pp. 67-100.

[2] F. Carlson, "Contributions à la théorie des séries de Dirichlet I," *Arkiv för Matematik Astronomi och Fysik*, vol. 16 (1922), No. 18.

[3] ———, "Contributions à la théorie des séries de Dirichlet II," *Arkiv för Matematik Astronomi och Fysik*, vol. 19 (1926), No. 25.

[4] G. Hardy, "*Ramanujan*," Cambridge University Press, 1940.

[5] B. Jessen and A. Wintner, "Distribution functions and the Riemann zeta function," *Transactions of the American Mathematical Society*, vol. 38 (1935), pp. 48-88.

[6] R. Rankin, "Contributions to the theory of Ramanujan's function $\tau(n)$ and similar arithmetical functions I," *Proceedings of the Cambridge Philosophical Society*, vol. 35 (1939), pp. 351-356.

[7] ———, "Contributions to the theory of Ramanujan's function $\tau(n)$ and similar arithmetical functions II," *Proceedings of the Cambridge Philosophical Society*, vol. 35 (1939), pp. 357-372.

[8] E. Study, "*Vorlesungen über ausgewählte Gegenstände der Geometrie*," B. G. Teubner, 1913.

[9] J. Wilton, "A note on Ramanujan's arithmetical function $\tau(n)$," *Proceedings of the Cambridge Philosophical Society*, vol. 25 (1929), pp. 121-129.

ANALYTIC FUNCTIONS OF BOUNDED TYPE.*

By A. Franck.

1. Introduction. An analytic function $f(z)$ is called *of bounded type* in a region if it can be represented as a quotient of two bounded analytic functions in that region, i. e. $f(z) = h_1(z)/h_2(z)$ where $|h_i(z)| < 1$, $(i = 1, 2)$. The following theorem is fundamental for the characterization of the class of functions of bounded type which are analytic in the unit circle.

Ostrowski-Nevanlinna Theorem.[1] (i) *A necessary and sufficient condition that a function $F(w)$, regular in the circle $|w| < 1$, be* B. T. *in $|w| < 1$ is that*

$$(1.1) \qquad \int_0^{2\pi} \log^+ |F(re^{i\theta})|\, d\theta \leqq M, \qquad 0 \leqq r < 1$$

where M is a constant.[2] (ii) *If $F(w)$ is* B. T. *in $|w| < 1$, then $F(w)$ may be represented in the form $F(w) = B(w)G(w)$ where $B(w)$ is the Blaschke product formed with the zeros $w_1, w_2, \cdots$ of $F(w)$ in $|w| < 1$, i. e. $B(w) = \prod_{k=1}^{\infty} \frac{w - w_k}{\bar{w}_k w - 1} \frac{\bar{w}_k}{|w_k|}$, the product being uniformly and absolutely convergent in every circle $|w| \leqq \rho < 1$; $G(w)$ is a non-vanishing function* B. T. *in the unit circle and $\log |G(w)|$ is representable by a Poisson-Stieltjes integral in $|w| < 1$*

$$(1.2) \qquad \log |G(w)| = \frac{1}{2\pi} \int_0^{2\pi} \frac{1 - \rho^2}{1 + \rho^2 - 2\rho \cos(\theta - \alpha)}\, d\phi(\theta),$$

where $w = \rho e^{i\alpha}$ and $\phi(\theta)$ is a function of bounded variation in $0 \leqq \theta \leqq 2\pi$. (iii) *If the inequality* (1.1) *is satisfied for $F(w)$, then it is also satisfied for $G(w)$ [with the same constant M].* (iv) *If $F(w)$ is* B. T. *in $|w| < 1$, it*

* Received March 8, 1950; revised May 4, 1951; presented to the American Mathematical Society, November 26, 1949. This paper contains a part of the results of the author's doctoral dissertation [1]. The author wishes to express his indebtedness to Professor S. E. Warschawski, under whose direction this work was done.

[1] Hereafter this theorem will be referred to as the O. N. Theorem. Furthermore a function of bounded type will be denoted by B. T.

[2] $\log^+ a = \log a$ for $a \geqq 1$ and $\log^+ a = 0$ for $a < 1$.

possesses boundary values for almost all θ, $(w = e^{i\theta})$ on the unit circle for non-tangential approach.

These results are due to A. Ostrowski [6]; condition (1. 1) and the result on the existence of the boundary values were also given in a joint paper by F. and R. Nevanlinna [4], p. 26.

The concept of a function of B. T. can be applied to a function which is meromorphic (rather than just regular) in a region. In a subsequent paper [5], R. Nevanlinna extended part (i) of the above theorem not only to functions meromorphic in the unit circle, but also to functions meromorphic in an arbitrary region. In the present paper we shall restrict ourselves to analytic functions.

The criteria obtained by R. Nevanlinna were based on the use of the Green's function for the regions considered. It would seem desirable to obtain criteria which are not based on a knowledge of the Green's function, which in general is difficult to obtain, but rather are expressed in terms of the geometry of the region of analyticity of the functions considered.

The purpose of the present paper is to establish such criteria for a fairly general class of regions. Applications of our methods permit us to derive several results which hold for special regions (such as the half-plane). In particular our criteria contain some of the results of A. Wishard [11] for the case of analytic functions.[3]

I. General criteria.

2. Extensions of the O. N. Theorem for the Half-Plane. The following lemma is an analogue of the O. N. Theorem stated for the half-plane.

LEMMA 1. *Suppose that $f(z)$ is analytic in $\mathrm{Re}(z) > 0$, $z = x + iy$. Let C_n be a sequence of circles, radii ρ_n, which exhaust[4] $x > 0$, $(n = 1, 2, \cdots)$. Suppose that for all n*

[3] Other classes of functions which are regular in a half-plane were treated by W. Kryloff [3]. He studies the classes of functions for which $(z = x + iy)$ (a) $\int_{-\infty}^{\infty} |f(z)|^p dy \leqq M^p < \infty$ and (b) $\int_{-\infty}^{\infty} \log^+ |f(z)|\, dy \leqq M < \infty$ and obtains "parametric representations" (in terms of certain Stieltjes integrals) for these functions.

[4] *A sequence of closed Jordan* (i. e. simple) *curves* C_n which are contained in a region R is said to exhaust the region R, if for any subregion R' whose closure $\bar{R'}$ is contained in R, there exists an integer N such that R' will be contained in the interior of C_n for all $n > N$. A similar definition is to be applied to the concept that *a sequence of regions exhaust a given region.*

$$\int_{C_n} \frac{\log^+ |f(z)|}{|1+z|^2} \rho_n \, d\theta \leqq M.^5 \tag{2.1}$$

Then $f(z)$ *is* B. T. *in* $x>0$; $f(z)=b(z)g(z)$ *where* (i) $b(z)$ *is the Blaschke product formed with the zeros* $z_1, z_2, \cdots$ *of* $f(z)$ *in* $x>0$; (ii) $g(z)\neq 0$; (iii) *for* $x>0$ *there exists a constant* λ *such that*

$$\log |g(z)| = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x \, d\mu(t)}{x^2+(y-t)^2} + \lambda x \tag{2.2}$$

where $\mu(t)$ *is a function of bounded variation in any finite interval and the integral converges absolutely*; (iv) *there exists a constant* N *such that for any circle* C *of radius* ρ *contained in* $x>0$

$$\int_{C} \frac{\log^+ |g(z)|}{|1+z|^2} \rho \, d\theta \leqq N. \tag{2.3}$$

Conversely if $f(z)$ *is* B. T. *in* $x>0$, *then it satisfies* (2.3) *with* f *replacing* g.

Proof. Our proof makes use of a theorem of R. Gabriel [2] for functions $U(w)$ non-negative and subharmonic in the unit circle. Let C_1 and C_2 *be circles contained in* $|w|<1$ *such that* C_2 *lies in the interior* of C_1. *Then*

$$\int_{C_2} U(w) |dw| \leqq 2 \int_{C_1} U(w) |dw|.$$

We map $x>0$ onto $|w|<1$ by means of the transformation

$$w = \frac{z-1}{z+1} \tag{2.4}$$

carrying the circles C_n onto a sequence of circles Γ_n which exhaust $|w|<1$. Points on C_n are given by $z=z_n+\rho_n e^{i\theta}$ where z_n is the center of C_n. Employing the inverse of (2.4) and $|dw| = 2\rho_n d\theta / |1+z|^2$, we find that

$$\int_{C_n} \frac{\log^+ |f(z)|}{|1+z|^2} \rho_n \, d\theta = \tfrac{1}{2} \int_{\Gamma_n} \log^+ |F(w)| \, |dw| \tag{2.5}$$

where $F(w)=f[(1+w)/(1-w)]$. Since the left side of (2.5) $\leqq M$. we have

$$\int_{\Gamma_n} \log^+ |F(w)| \, |dw| \leqq 2M. \tag{2.6}$$

Consider the circle $|w|=r_0<1$. Keeping r_0 fixed, choose n so large that Γ_n will contain this circle in its interior. Applying Gabriel's Theorem

[5] As a matter of convention the letters $M, M_1, M_2, \cdots, N, N_1, N_2, \cdots$ will always designate constants.

to $\log^+ |F(w)|$, we find from (2.6) that $\int_0^{2\pi} \log^+ |F(r_0 e^{i\phi})|\, d\phi \leqq 4M/r_0$. Since $\int_0^{2\pi} \log^+ |F(re^{i\phi})|\, d\phi$ is a monotonic increasing function of r, we have for any circle $|w| = r < r_0$: $\int_0^{2\pi} \log^+ |F(re^{i\phi})|\, d\phi \leqq 4M/r_0$. Keeping r fixed, let $r_0 \to 1$. Thus $\int_0^{2\pi} \log^+ |F(re^{i\phi})|\, d\phi \leqq 4M$ for all r, $0 \leqq r < 1$. Hence by the O. N. Theorem $F(w)$ is B. T. in $|w| < 1$; $F(w) = B(w)G(w)$ where $G(w) \neq 0$. Returning to $x > 0$, we find $f(z) = b(z)g(z)$ where $g(z) = G[(z-1)/(z+1)] \neq 0$. By means of the mapping (2.4) and (1.2) of the O. N. Theorem, the representation for $\log |g(z)|$ is established.

To establish (2.3) we map $x > 0$ onto $|w| < 1$ carrying a circle C in $x > 0$ into a circle Γ in $|w| < 1$. Let r_0 be the maximal distance of Γ from $w = 0$. Construct any circle $|w| = r > r_0$, $0 \leqq r < 1$. Applying Gabriel's Theorem to $\log^+ |G(w)|$ we have

$$\int_\Gamma \log^+ |G(w)|\, |dw| \leqq 2 \int_0^{2\pi} \log^+ |G(re^{i\phi})|\, r\, d\phi. \tag{2.7}$$

From the O. N. Theorem the integral on the right side of (2.7) is bounded for all r, $0 \leqq r < 1$. Transforming the left side of (2.7) by (2.4), we establish (2.3).

To prove the converse merely repeat for $f(z) = f[(1+w)/(1-w)] = F(w)$ the argument which we applied to $g(z) = G(w)$ in the preceding paragraph.

The following theorem is readily established from Lemma 1.

THEOREM I. *Let $f(z)$ be analytic in $x > 0$. Let C_n be a sequence of circles, radii R_n, which exhaust $x > 0$. Suppose that for all n*

$$\iint_{C_n} \frac{\log^+ |f(z)|}{|1+z|^2}\, dx\, dy \leqq MR_n. \tag{2.8}$$

Then $f(z)$ is B. T. *in $x > 0$; $f(z) = b(z)g(z)$ and for any circle C in $x > 0$*

$$\iint_C \frac{\log^+ |g(z)|}{|1+z|^2}\, dx\, dy \leqq NR, \qquad (R = \text{radius of } C). \tag{2.9}$$

Conversely, if $f(z)$ is B. T. *in $x > 0$. then $f(z)$ satisfies* (2.9) *with f replacing g.*

Proof. Consider the annular rings A_{R_n}: $(R_n/2 \leqq |z - z_n| \leqq R_n)$ where z_n is the center of C_n. Employing polar coordinates about $z = z_n$, we find from (2.8) that

$$\int_{R_n/2}^{R_n}\left[\int_0^{2\pi}\frac{\log^+|f(z)|}{|1+z|^2}d\theta\right]r\,dr \leqq MR_n. \tag{2.10}$$

Applying the mean value theorem to the outer integral of (2.10), we have

$$\iint_{A_{R_n}}\frac{\log^+|f(z)|}{|1+z|^2}d\theta\, r\,dr = \frac{3R^2_n}{8}\int_0^{2\pi}\frac{\log^+|f(\zeta)|}{|1+\zeta|^2}d\theta \tag{2.11}$$

where $\zeta = z_n + \rho_n e^{i\theta}$ and $R_n/2 < \rho_n < R_n$.

Since $\rho_n < R_n$, we have from (2.10) and (2.11)

$$\int_0^{2\pi}\frac{\log^+|f(\zeta)|}{|1+\zeta|^2}d\theta \leqq \frac{8M}{3\rho_n}.$$

Thus the conditions of Lemma 1 are satisfied for the circles $|z - z_n| = \rho_n$ and hence $f(z)$ is B. T. in $x > 0$. From Lemma 1 it follows that for *any* circle C in $x > 0$, we have $\int_C \frac{\log^+|g(z)|}{|1+z|^2}\rho\, d\theta \leqq N$. From this, inequality (2.9) follows. The converse also follows from the converse of Lemma 1.

3. An area theorem for arbitrary regions. In the present section we shall extend our results to obtain conditions under which an analytic function will be B. T. in a region [6] of a more general nature than a circle or a half-plane. The following theorem which is an extension of Theorem I gives a sufficient condition for an analytic function to be B. T. in a region R.

THEOREM II. (a) *Let $f(z)$ be analytic in a bounded region R. Let R_n be a sequence of regions which exhaust R. Let $w = \phi(z)$ map R conformally onto $|w| < 1$ so that a point $z = a$ on the boundary of R correspond to $w = 1$, suppose that* (b) *$|\phi'(z)| \leqq m$ in R where m is a constant and* (c) $\left|\frac{1-w}{a-z}\right| \geqq N > 0$ *in R, where $w = \phi(z)$.*[7] (d) Let $d_n = \sup d_n(z)$ *for all z on the boundary of R_n where $d_n(z)$ is the minimal distance of the point z on the boundary of R_n to the boundary of R.*

---

[6] By region we mean a simply-connected region. This usage will be applied throughout the remainder of the paper.

[7] Hypotheses (b) and (c) of this theorem will, for example, be satisfied if R is bounded by a closed Jordan curve C, with the following properties: (i) one can roll a circle on the exterior of C (i. e. to each point ζ of C, one can associate a circle passing through ζ such that its interior lies in the exterior of C, each circle having the same radius); (ii) at the point $z = a$, the curve possesses finite curvature (by this we mean that C possesses a tangent at every point of a neighborhood of $z = a$ and that the curvature exists and is finite at $z = a$).

Conditions (i) implies (b), for example, see [9]; condition (ii) implies (c), for example, see [10], p. 433 or [8], Theorems 18 and 19.

Then $f(z)$ *is* B. T. *in* R *if*

$$\int\int_{R_n} \frac{\log^+ |f(z)|}{|a-z|^2}\, dx\, dy \leqq \frac{M}{d_n} \tag{3.1}$$

Proof. Since $w=\phi(z)$ maps R onto $|w|<1$, it also maps the regions R_n onto regions W_n contained in $|w|<1$ which exhaust the unit circle. If n is taken sufficiently large, the regions W_n will contain the point $w=0$. Consider such regions which contain $w=0$ and let r_n be the minimal distance of the boundary of W_n from $w=0$. Let us consider a point on the boundary of W_n for which the minimal distance is assumed. Designate it as $w=r_ne^{i\alpha_n}$. Let z_n be the image point in the z-plane as defined by $r_ne^{i\alpha_n}=\phi(z_n)$. Consider the integral $\int_{z_n}^{z_0} |\phi'(z)|\,|dz|$ along the straight line segment joining z_n to the nearest point z_0 on the boundary of R. Transforming this integral to the w-plane by means of the mapping function, we have

$$\int_{z_n}^{z_0} |\phi'(z)|\,|dz| = \int_{r_ne^{i\alpha_n}}^{e^{i\beta}} |dw|, \qquad [e^{i\beta}=\phi(z_0)]; \tag{3.2}$$

the last integral being taken over the image curve of the line segment from z_n to z_0. From hypothesis (b) and the fact that the length of any curve from any point w in $|w|<1$ to the boundary is greater than or equal to $1-|w|$, we find from (3.2) that $md_n(z_n)\geqq 1-r_n$ where $d_n(z_n)$ is the minimal distance from the point z_n to the boundary of R. Thus, we finally have that

$$d_n \geqq (1-r_n)/m. \tag{3.3}$$

Now let $F(w)=f[\psi(w)]$ where $z=\psi(w)$ is the inverse of $w=\phi(z)$. On transforming the integral in (3.1) to the w-plane. we find from (b) and (c) that

$$\int\int_{R_n} \frac{\log^+|f(z)|}{|a-z|^2}\, dx\, dy = \int\int_{W_n} \frac{\log^+|F(w)|}{|1-w|^2}\ \frac{|1-w|^2}{|a-z|^2}\, |\psi'(w)|^2\, du\, dv$$
$$\geqq \frac{N^2}{m^2}\int\int_{W_n} \frac{\log^+|F(w)|}{|1-w|^2}\, du\, dv$$

as $|\psi'(w)|=1/|\phi'(z)|\geqq 1/m$. Taking note of (3.1) and (3.3) we find that

$$\int\int_{W_n} \log^+|F(w)|/|1-w|^2\, du\, dv \leqq (m^3M/N^2)/(1-r_n).$$

By means of the mapping $w=(\zeta-1)/(\zeta+1)$ to $\mathrm{Re}(\zeta)\geqq 0$, one can show that condition (2.8) is satisfied. Hence it follows by Theorem I that $f(z)=F[\phi(z)]$ is B. T. in R.

As a converse to this theorem we establish the following corollary.

COROLLARY. *Let $f(z)$ be* B. T. *in a bounded region R. Let $w = \phi(z)$ map the region R onto the unit circle $|w| < 1$ in such a manner that a point $z = a$ on the boundary of R corresponds to $w = 1$. Suppose that $|\phi'(z)| \geqq m_1 > 0$ in R where m_1 is a constant and that $\left|\frac{1-w}{a-z}\right| \leqq N_1$ in R where $w = \phi(z)$.*

Then there exists a sequence of regions R_n which exhaust R such that

$$\iint_{R_n} [\log^+ |f(z)|/|a-z|^2]\,dx\,dy \leqq N/d_n \tag{3.4}$$

where d_n has the same meaning as in Theorem II.

Proof. Consider a sequence of circles Γ_n in $|w| < 1$ which exhaust the unit circle. By employing an argument similar to that of the proof of Theorem II, one can show that the images, R_n, of the interior of the Γ_n form a sequence for which (3.4) is satisfied.

REMARK. *If the region R is such that there exists two positive constants m_1 and m_2 for which $m_1 \leqq |\phi'(z)| \leqq m_2$,[8] then condition* (3.1) *is necessary and sufficient for $f(z)$ to be* B. T. *in R.*

The following theorem can also be deduced from the O. N. Theorem. It is merely stated here. For a proof of this results, see [1].

THEOREM III. *Let $f(z)$ be analytic in the interior R of a closed Jordan curve. Suppose that $w = \phi(z)$ maps R conformally onto $|w| < 1$ so that $0 < m_1 \leqq |\phi'(z)| \leqq m_2$[8] for all z in R where m_1 and m_2 are constants. Then a necessary and sufficient condition for $f(z)$ to be* B. T. *in R is that there exist some sequences of regions R_n which exhaust R, for which*

$$\iint_{R_n} [\log^+ |f(z)|/d(z)]\,dx\,dy \leqq M \log 1/d_n$$

where $d(z)$ is the distance of the point z to the boundary of R and d_n has the same meaning as in Theorem II.

II. Applications.

4. A criterion for a circular sector. As an application of the general area Theorem II, we establish

THEOREM IV. *Let $f(z)$ be analytic in the sector*

$$S: (0 < r < 1, -\alpha < \theta < \alpha).$$

[8] This condition can be characterized in terms of the geometry of the boundary of R (see [8], Theorems 18 and 19 and [10], p. 433).

Suppose that $\int_{-a}^{a}\log^+|f(\rho e^{i\theta})|\,d\theta \leqq M$, $0<\rho<1$. *Then* $f(z)$ *is* B. T. *in the sector* S.[9]

Proof. Let S_n be a sequence of "truncated" sectors

$$S_n[1-r_n \leqq \rho \leqq r_n < 1; -\alpha_n \leqq \theta = \arg z \leqq \alpha_n = \alpha-(1-r_n)]$$

where, for $n=1,2,\cdots$, $0<r_n\leqq r_{n+1}<1$, $1-r_n<\alpha$, and $\lim_{n\to\infty} r_n = 1$. The sequence S_n clearly exhaust S. We note that $d_n = 1-r_n$ (see Theorem II.). Furthermore, conditions (a), (b) and (c) of Theorem II are satisfied where the point $a=1$ (see footnote 7). Now

$$\iint_{S_n}\frac{\log^+|f(z)|}{|1-z|^2}\,dxdy = \int_{1-r_n}^{r_n}\int_{-\alpha_n}^{\alpha_n}\frac{\log^+|f(\rho e^{i\theta})|\rho\,d\rho\,d\theta}{|1-\rho e^{i\theta}|^2}$$

$$\leqq \int_{1-r_n}^{r_n}\frac{\rho}{(1-\rho)^2}\left[\int_{-\alpha_n}^{\alpha_n}\log^+|f(\rho e^{i\theta})|\,d\theta\right]d\rho \leqq M\int_{1-r_n}^{r_n}\frac{\rho}{(1-\rho)^2}\,d\rho$$

the last inequality following by hypothesis. We now have

$$\iint_{S_n}\log^+|f(z)|/|1-z|^2\,dx\,dy \leqq M\int_0^{r_n}d\rho/(1-\rho)^2 \leqq M/(1-r_n) = M/d_n.$$

Thus, condition (3.1) of Theorem II is satisfied and hence $f(z)$ is B. T. in S.

5. **Some criteria for the half-plane.** In the present section we shall give two applications of Theorem I which are concerned with stating necessary and sufficient conditions for an analytic function to be B. T. in $x>0$. As a first example we shall establish the following theorem.

THEOREM V. *Let* $f(z)$ *be analytic in* $x>0$. *Suppose that for all* θ, $-\pi/2<\theta<\pi/2$,

$$\int_{1/R}^{R}\log^+|f(z)|/|1+z|^2\,r\,dr \leqq MR, \qquad (z=re^{i\theta}). \tag{5.1}$$

[9] This theorem is somewhat similar to a theorem due to I. Privaloff [7] for functions subharmonic in a sector, i. e. *If* $v(z)$ *is a subharmonic function in the open sector* $\alpha<\arg z<\beta$, $|z|<1$ *and if* $\int_\alpha^\beta v^+(re^{i\theta})\,d\theta = O(1)$, *then at almost any point of the arc* (α,β) *of the unit circle the limit of* $v(z)$ *exists along any path not touching the arc.* $[v^+(z)=v(z)$ *if* $v(z)\geqq 0$; $v^+(z)=0$ *if* $v(z)<0]$.

Then $f(z)$ *is* B. T. *in* $x > 0$ *and* $f(z) = b(z)g(z)$. *Furthermore for all* θ, $-\pi/2 < \theta < \pi/2$,

$$\int_{1/R}^{R} \log^+ |\, g(z)\,|/|\, 1 + z\,|^2\, r\, dr \leqq NR. \tag{5.2}$$

Conversely, if $f(z)$ *is* B. T. *in* $x > 0$, *then* (5.2) *holds with* g *replaced by* f.

Proof. We integrate (5.1) with respect to θ and find

$$\int_{-\pi/2}^{\pi/2} \int_{1/R}^{R} \log^+ |\, f(z)\,|/|\, 1 + z\,|^2\, r\, dr\, d\theta \leqq \pi MR.$$

Consider the circles C, in the semi-circular ring A: $(-\pi/2 \leqq \theta \leqq \pi/2$, $1/R \leqq |\, z\,| \leqq R)$ which have their centers on the real axis and pass through the points $x = 1/R$ and $x = R$. The radii of these circles are $\bar{R} = \dfrac{(R + 1/R)}{2}$. We see that $R \leqq 2\bar{R}$. Thus

$$\int\int_C \log^+ |\, f(z)\,|/|\, 1 + z\,|^2\, dx\, dy \leqq 2\pi M\bar{R}.$$

Since the circles C exhaust $x > 0$, the conditions of Theorem I are fulfilled. Hence $f(z)$ is B. T. in $x > 0$.

To establish the condition on $g(z)$ assuming $f(z)$ is B. T., we proceed as follows. From the representation for $\log |\, g(z)\,|$ given by (2.2) we see that

$$\log^+ |\, g(z)\,| \leqq \frac{1}{\pi} \int_{-\infty}^{\infty} [x^2 + (y - t)^2]^{-1} x\ |\, d\mu(t)\,| + |\,\lambda\,|\, x. \tag{5.3}$$

Multiplying both sides of (5.3) by $r/(1 + r^2)$ and integrating with respect to r, we find

$$\begin{aligned} &\int_{1/R}^{R} \frac{\log^+ |\, g(z)\,|}{1 + r^2}\, r\, dr \\ &\leqq \frac{1}{\pi} \int_{-\infty}^{\infty} \left[\int_{1/R}^{R} \frac{xr}{x^2 + (y - t)^2} \cdot \frac{dr}{1 + r^2} \right] |\, d\mu(t)\,| + |\,\lambda\,| \int_{1/R}^{R} \frac{xr\, dr}{1 + r^2}. \end{aligned} \tag{5.4}$$

Let us consider the integral in the brackets of (5.4), i. e.

$$I = \int_{1/R}^{R} \frac{xr}{x^2 + (y - t)^2} \cdot \frac{dr}{1 + r^2}. \tag{5.5}$$

To estimate this integral we consider two cases:

(a) $0 < |t| < 2R$. We employ contour integration to estimate I. It is seen that

$$I \leqq \operatorname{Re}\left[\int_{-\infty}^{\infty} \frac{r}{z - it} \frac{dr}{1 + r^2}\right]. \tag{5.6}$$

Let us consider the complex integral $\displaystyle\int \frac{\zeta}{\zeta e^{i\theta} - it} \frac{d\zeta}{1 + \zeta^2}$, the contour over which the integration takes place being specified below.

(i) For $t > 0$ we consider a contour in $\operatorname{Im}(\zeta) \leqq 0$ defined by the semicircle C_R: $(|\zeta| = R)$ and the line segment $-R \leqq \zeta \leqq R$. We thus have

$$\int_{1/R}^{R} \frac{\zeta}{\zeta e^{i\theta} - it} \frac{d\zeta}{1 + \zeta^2} + \int_{C_R} \frac{\zeta}{\zeta e^{i\theta} - it} \frac{d\zeta}{1 + \zeta^2} = -2\pi i S$$

where S is the residue of the integrand of the complex integral at its only pole $\zeta = -i$ in $\operatorname{Im}(\zeta) \leqq 0$. Thus $\displaystyle\int_{-R}^{R} = \pi/(e^{i\theta} + t) - \int_{C_R}$. Letting $R \to \infty$ and noting that the integral over C_R vanishes, we have

$$\int_{-\infty}^{\infty} \frac{\zeta}{\zeta e^{i\theta} - it} \frac{d\zeta}{1 + \zeta^2} = \frac{\pi}{e^{i\theta} + t}. \tag{5.7}$$

Taking the real part of both sides of (5.7) we have from (5.6) that

$$I \leqq \pi(\cos\theta + t)/(1 + t^2 + 2t\cos\theta)$$

As $t > 0$ and $-\pi/2 < \theta < \pi/2$, we see that $t\cos\theta \geqq 0$. Further as $|t| < 2R$, we find

$$I \leqq N_1 R/(1 + t^2) \tag{5.8}$$

(ii) For $t < 0$ we consider a contour in $\operatorname{Im}(\zeta) \geqq 0$ defined by the semicircle C_R: $(|\zeta| = R)$ and the line segment $-R \leqq \zeta \leqq R$. An argument similar to that given in (i) will show that

$$I \leqq N_2 R/(1 + t^2) \tag{5.9}$$

(b) $|t| \geqq 2R > 1$. We return to (5.5). Using polar coordinates we find

$$I = \cos\theta \int_{1/R}^{R} \frac{r^2}{r^2 + t^2 - 2rt\sin\theta} \frac{dr}{1 + r^2}, \tag{5.10}$$

Now $(r^2 + t^2 - 2rt \sin \theta) \geqq (|t| - r)^2 \geqq t^2/4$. Since $1/t^2 \leqq 2/(1 + t^2)$, we obtain on estimating (5.10)

$$I \leqq 8R/(1 + t^2). \tag{5.11}$$

Since the second integral of (5.4) $\leqq R$, we have from (5.8), (5.9), and (5.11) by splitting the integral of (5.4) into the cases discussed that

$$\int_{1/R}^{R} \log^+ |g(z)| \, r/(1 + r^2) dr \leqq N_8 R \int_{-\infty}^{\infty} | d\mu(t) | /(1 + t^2) + |\lambda| R. \tag{5.12}$$

But the integral on the right side of (5.12) is convergent. Hence the result is established. The converse of the theorem follows from this result.

As a second example we present a new and rather simple proof of a theorem due to A. Wishard [11], Theorem C, p. 672. This theorem was originally stated for a function meromorphic in a half plane. In this paper. however, we shall only establish this theorem for analytic functions.

THEOREM VI. *Let $f(z)$ be analytic in $x > 0$. Suppose that for all $x > 0$*

$$\int_{-\infty}^{\infty} \log^+ |f(z)| / |1 + z|^2 \, dy \leqq M, \qquad (z = x + iy). \tag{5.13}$$

Then $f(z)$ is B. T. *in $x > 0$; $f(z) = b(z) g(z)$ and for all $x > 0$*

$$\int_{-\infty}^{\infty} \log^+ |g(z)| / |1 + z|^2 \, dy \leqq N. \tag{5.14}$$

Conversely, if $f(z)$ is B. T. *in $x > 0$, then* (5.14) *holds with g replaced by f.*

Proof. Let us consider the sequences of circles C_n, radii R_n, whose centers $a_n = R_n + \delta_n$ lie on the real-axis and $\delta_n > 0$. If $\delta_n \to 0$ and $R_n \to \infty$, this sequence will exhaust $x > 0$. Noting that the integral of

$$(\log^+ f(z)) / |1 + z|^2$$

taken over the area of the circle C_n is less than that over the circumscribed square, rewriting the integral over C_n as an iterated integral and employing (5.13) we find

$$\int_{C_n} \log^+ |f(z)| / |1 + z|^2 \, dx \, dy \leqq 2MR_n.$$

Hence by Theorem I, it follows that $f(z)$ is B. T. in $x > 0$.

By means of (5.3) we shall establish (5.14). We find that

$$(5.15) \qquad \int_{-T_1}^{T_2} \frac{\log^+ |g(z)|}{|1+z|^2}\, dy$$
$$\leqq \frac{1}{\pi} \int_{-T_1}^{T_2} \int_{-\infty}^{\infty} \frac{x\,|d\mu(t)|}{x^2+(y-t)^2} \frac{dy}{|1+z|^2} + |\lambda| \int_{-T_1}^{T_2} \frac{x\,dy}{|1+z|^2}.$$

Interchanging the order of integration in (5.15) and noting that

$$\frac{1}{1+x} \cdot \frac{2x+1}{(1+2x)^2+t^2} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{1}{(1+x)^2+y^2} \cdot \frac{x\,dy}{x^2+(y-t)^2}$$

we find

$$(5.16) \qquad \int_{-T_1}^{T_2} \frac{\log^+ |g(z)|}{|1+z|^2}\, dy$$
$$\leqq \frac{2x+1}{x+1} \int_{-\infty}^{\infty} \frac{|d\mu(t)|}{(1+2x)^2+t^2} + |\lambda| \int_{-\infty}^{\infty} \frac{x\,dy}{|1+z|^2}.$$

We note from Lemma I that the integral $\int_{-\infty}^{\infty} |d\mu(t)|/(1+t^2)$ is convergent. Furthermore $\int_{-\infty}^{\infty} dy/|1+z|^2 = \pi/(1+x)$. Thus, for all $x>0$ and any T_1 and T_2,

$$\int_{-T_1}^{T_2} \log^+ |g(z)|/|1+z|^2\, dy \leqq N.$$

Now let $T_1 \to \infty$ and $T_2 \to \infty$. Hence (5.14) is established. The converse follows easily from this result. For if $f(z)$ is B. T., then $|f| = |b|\,|g| < |g|$; but $\log^+ |g(z)|$ satisfies (5.3) and hence the conclusion follows.[10]

UNIVERSITY OF MINNESOTA.

---

[10] The O. N. Theorem states that if $F(w)$ is analytic in $|w|<1$ and

$$\int_0^{2\pi} \log^+ |F(re^{i\theta})|\, d\theta \leqq M$$

for all r, $0 \leqq r < 1$, then $F(w) = B(w)G(w)$ and

$$\int_0^{2\pi} \log^+ |G(re^{i\theta})|\, d\theta \leqq M$$

for $0 \leqq r < 1$ with *exactly the same constant M*. If Theorem VI is restated for the unit circle, it is seen that the vertical lines correspond to a sequence of circles passing through the point $w=1$ which exhaust the unit circle. From this point of view it would be natural to ask whether the constants M and N in (5.13) and (5.14) are the same also. An example can be given for which $M < N$ (see [1]).

BIBLIOGRAPHY

[1] A. Franck,, On analytic functions of bounded type, Doctoral Thesis, University of Minnesota, June 1949.

[2] R. M. Gabriel, "An inequality concerning the integrals of positive subharmonic functions along certain curves," *Journal of the London Mathematical Society*, vol. 4 (1929), pp. 133-139.

[3] W. Kryloff, "Uber Funktionen, die in der Halbebene regulär sind," *Mat. Sbornik*, new series volume 6 [old 48] (1939), pp. 95-138.

[4] F. and R. Nevanlinna, "Über die Eigenschaften analytischer Funktionen in der Umgebung einer singulären Stelle oder Linie," *Acta Soc. Sc. Fennicae*, vol. 50 (1922), No. 5.

[5] R. Nevanlinna, "über eine Klasse meromorpher Funktionen," *Mathematische Annalen*, vol. 92 (1924), pp. 145-154.

[6] A. Ostrowski, "Über die Bedeutung der Jensenschen Formel fur einige Fragen der komplexen Funktiontheorie," *Acta Litt. ac Scient. Regiae Univ.* Hung. Francisco-Josephinae, vol. 1 (1923), pp. 80-87.

[7] I. Privaloff, "Sur un problème limite de la théorie des fonctions," *Bull. Acad. Sci. USSR*, vol. 7 (1935), pp. 301-304.

[8] W. Seidel, "Über die Ränderzuordnung bei konformen Abbildungen," *Mathematische Annalen*, vol. 104 (1931), pp. 182-242.

[9] C. Visser, "De conforme afbeelding van een speciale klasse van gebieden," *Handelingen van het XXVe Nederlandsch Natuur-en Geneeskundig Congress*, Delft, 1935.

[10] S. E. Warschawski, "Uber das Randverhalten der Ableitung der Abbilddungsfunktion bei konformer Abbildung," *Mathematische Zeitshrift*, vol. 35 (1932), pp. 322-456.

[11] A. Wishard, "Functions of bounded type," *Duke Mathematical Journal*, vol. 9 (1942), pp. 663-676.

INTEGRAL GEOMETRY IN HERMITIAN SPACES.*

By L. A. Santaló.

1. Introduction. Let P_n be the n-dimensional complex projective space with the homogeneous coordinates x^i $(i = 0, 1, 2, \cdots, n)$. Let x^{*i} denote the complex conjugate of x^i.

We consider the group U (unitary group) of linear transformations

$$x'^i = \sum_{k=0}^{n} \xi_k{}^i x^k \tag{1.1}$$

which leaves the form

$$(xx^*) = \sum_{i=0}^{n} x^i x^{*i} \tag{1.2}$$

invariant. Here and in the sequel, a^* denotes the complex conjugate $(=\bar{a})$ of a. The coefficients of $\xi_k{}^i$ satisfy, then, the relations

$$(\xi_k \xi^*_l) = \sum_{i=0}^{n} \xi_k{}^i \xi^*{}_l{}^i = \delta_{kl}, \tag{1.3}$$

where $\delta_{kl} = 0$ if $k \neq l$ and $= 1$ if $k = l$.

Since the coefficients $\xi_k{}^i$ are complex numbers, the group U depends upon $(n+1)^2$ real parameters.

The geometry of P_n with the fundamental group U of transformations is called the Hermitian geometry (more precisely the elliptic hermitian geometry) and the space P_n itself is called an Hermitian space. The Integral Geometry in these spaces was initiated by Blaschke [2] who gave the densities for linear subspaces, without making applications to integral formulas. The case $n = 2$ was first considered by Varga [8] and later, in a more complete form, by Rohde [6].

In the present paper we generalize to the n-dimensional case some of the results which Varga and Rohde obtained for the plane. The main results we obtain are the following:

We first determine the explicit form of the left invariant element of volume du of the group U. If $L_r{}^0$ is a fixed linear subspace of dimension r (throughout the paper we shall mean by *dimension* the complex dimension)

* Received August 27, 1951.

and Γ_r is the subgroup of U which leaves L_r^0 invariant, we determine also the explicit form of the invariant element of volume dL_r of the homogeneous space U/Γ_r. Then, an easy calculation gives the total volumes (6.3) and (6.4) of U and U/Γ_r.

If the coordinates x^i are normalized such that $(xx^*) = 1$ it is well known that every variety C_p of dimension p $(p < n)$ has an invariant integral of degree $2p$, namely $J_p(C_p)$ defined in section 4. Let C_h and C_r $(r + h \geq n)$ be two analytic varieties. Let uC_r denote the transform of C_r by the transformation u of U, and let $C_h \cap uC_r$ denote the $(h + r - n)$-dimensional variety intersection of C_h and uC_r. If the invariant element of volume du of U is normalized in such a way that the total volume of U is equal 1, we prove the following integral formula:

$$\int_U J_{h+r-n}(C_h \cap uC_r)\, du = J_h(C_h) J_r(C_r). \tag{1.4}$$

If instead of C_r we consider a linear subspace L_r, and the invariant element of volume dL_r in U/Γ_r is normalized such that the total volume of U/Γ_r is equal 1, we also have

$$\int_{U/\Gamma_r} J_{h+r-n}(C_h \cap L_r)\, dL_r = J_h(C_h). \tag{1.5}$$

We do not consider the case in which C_h is non-analytic. For $n = 2$ this more general case has been considered by Rohde [6], and it remains an interesting open question to extend the results of Rohde to the n-dimensional case.

2. Relative components and equations of structure of the unitary group. The coefficients ξ_k^i of the unitary transformation (1.1) can be interpreted as coordinates of the transformed points ξ_k $(k = 0, 1, \cdots, n)$ of the $n + 1$ vertices of the n-simplex of reference; according to (1.3) they will be the vertices of an autoconjugate n-simplex with respect to the fundamental quadric $(\xi\xi^*) = 0$. These autoconjugate n-simplexes may be considered as the "frames" for the unitary group U (according to the theory of Cartan [3]) and the relative components ω_{jk} will then be defined by the equations

$$d\xi_k = \sum_{j=0}^{n} \omega_{kj}\xi_j. \tag{2.1}$$

From (2.1) and (1.3) de deduce

$$\omega_{kj} = (\xi^*_j d\xi_k), \tag{2.2}$$

and the equations of structure, deduced from (2.2) by remembering that $d(d\xi_k) = 0$, are

$$d\omega_{ih} = \sum_{k=0}^{n} [\omega_{ik}\omega_{kh}], \tag{2.3}$$

where here and throughout the paper square brackets denote exterior multiplication [3].

Notice that if ω^*_{kj} denotes the complex conjugate of the pfaffian form ω_{kj}, from (1.3) and (2.2) we deduce

$$\omega_{kj} + \omega^*_{jk} = 0. \tag{2.4}$$

3. Density for linear subspaces and cinematic density. We wish to define a measure for sets of r-dimensional linear subspaces L_r invariant with respect to U. If L_r^0 is a fixed L_r, and Γ_r denotes the subgroup of U which leaves L_r^0 invariant, then the problem is equivalent to the determination of an invariant element of volume in the homogeneous space U/Γ_r.

We follow the general method [5], [7]. Let L_r^0 be defined by the points $\xi_0, \xi_1, \cdots, \xi_r$. If it is fixed, in the equations (2.1) we will have

$$\omega_{jk} = 0 \quad \text{for} \quad 0 \leq j \leq r, \quad r+1 \leq k \leq n. \tag{3.1}$$

Since ω_{jk} are complex pfaffian forms, from $\omega_{jk} = 0$ we deduce $\omega^*_{jk} = 0$. Consequently the number of forms ω_{jk}, ω^*_{jk} $(0 \leq j \leq r,\ r+1 \leq k \leq n)$ is $2(r+1)(n-r)$. The density for sets of L_r, that is, the invariant element of volume in the space U/Γ_r will be, up to a constant factor,

$$dL_r = [\prod \omega_{jk}\omega^*_{jk}], \tag{3.2}$$

where the exterior product is taken always in absolute value and the indices range between the limits

$$0 \leq j \leq r, \qquad r+1 \leq k \leq n. \tag{3.3}$$

According to the equations of structure (2.3), it is easy to verify that $d(dL_r) = 0$, which is a sufficient condition in order that dL_r be effectively a density [7].

The invariant element of volume du in the space of the group U ("cinematic density" in the nomenclature of Blaschke) will be, up to a constant factor,

$$du = [\prod \omega_{jk}\omega^*_{jk} \prod \omega_{hh}] \tag{3.4}$$

where $j < k$, $0 \leq j, k, h \leq n$, that is, according to (2.4), du is equal to the absolute value of the exterior product of all the relative components.

4. The invariant integral $J_r(C_r)$ of an analytic variety C_r. Let C_r be an analytic variety of complex dimension r, that is, a variety defined by a set of $n+1$ analytic functions $x^i = x^i(t_1, t_2, t_3, \cdots, t_r)$ $(i = 0, 1, 2, \cdots, n)$ of r complex variables $t_1, t_2, \cdots, t_r$ in a domain D.

Assuming the homogeneous coordinates x^i normalized such that

$$(xx^*) = \sum_{i=0}^{n} x^i x^{*i} = 1, \tag{4.1}$$

let us consider the following differential form of degree $2r$:

$$\Omega^r = \sum [dx^{i_1} dx^{*i_1} \cdots dx^{i_r} dx^{*i_r}], \tag{4.2}$$

the summation being extended over all the combinations of $i_1, i_2, \cdots, i_r$ from 1 to n.

It is well known [4] that Ω^r is the only differential form of degree $2r$ which is invariant with respect to the group U.

The integral of Ω^r over an r-dimensional linear subspace L_r has the value [4],

$$\int_{L_r} \Omega^r = (2\pi i)^r / r!. \tag{4.3}$$

For a general analytic variety C_r we introduce the invariant integral $J_r(C_r)$ defined by

$$J_r(C_r) = r!/(2\pi i)^r \int_{C_r} \Omega^r. \tag{4.4}$$

If C_r is an algebraic variety, $J_r(C_r)$ coincides with its order (Cartan [4]).

For some purpose it is convenient to write Ω^r in another form. Let T_r be the r-dimensional linear subspace tangent to C_r at the point x, and take in T_r r points α_p $(p = 1, 2, \cdots, r)$ such that

$$(x\alpha^*_p) = 0, \qquad (\alpha_p \alpha^*_q) = \delta_{pq}. \tag{4.5}$$

We will have

$$dx^i = ax^i + \sum_{p=1}^{r} b_p \alpha_p{}^i, \qquad dx^{*i} = a^* x^{*i} + \sum_{p=1}^{r} b^*{}_p \alpha^*{}_p{}^i \tag{4.6}$$

where a and b_p are pfaffian forms given by

$$a = (x^* dx), \qquad b_p = (\alpha^*{}_p dx). \tag{4.7}$$

From (4.6) and (4.5) we deduce

$$\sum_{i=0}^{n} [dx^i dx^{*i}] = [aa^*] + \sum_{p=1}^{r} [b_p b^*{}_p]. \tag{4.8}$$

Since $a = -a^*$, we have $[aa^*] = 0$, and taking the r-th power of both sides of (4.8) we get $\Omega^r = [b_1 b^*_1 b_2 b^*_2 \cdots b_r b^*_r]$, or, according to (4.7),

$$\Omega^r = [\prod_{p=1}^{r} (\alpha_p{}^* dx)(\alpha_p dx^*)], \tag{4.9}$$

a formula which will be useful in the following sections.

5. The invariant integral J_r for the variety generated by the $(r-1)$-osculating spaces of a given analytic curve. As an example, we wish to evaluate J_r for the variety generated by the $(r-1)$-osculating linear spaces of a given analytic curve of the n-dimensional complex projective space.

Let C be an analytic curve defined by the $n+1$ parametric equations $y^i = y^i(t)$ $(i = 0, 1, 2, \cdots, n)$, where $y^i(t)$ are analytic functions of the complex variable t. In order to evaluate $J_1(C)$ we must normalize the coordinates y^i as indicated in (4.1). We set $x^i = y^i/(yy^*)^{\frac{1}{2}}$; then

$$dx^i = \{2(yy^*)dy^i - (dyy^*)y^i - (ydy^*)y^i\}/2(yy^*)^{3/2},$$
$$dx^{*i} = \{2(yy^*)dy^{*i} - (dy^*y)y^{*i} - (y^*dy)y^{*i}\}/2(yy^*)^{3/2},$$

and by exterior multiplication and addition,

$$\sum_{i=0}^{n} [dx^i dx^{*i}] = \{(yy^*)\sum [dy^i dy^{*i}] - [(y^*dy)(ydy^*)]\}/(yy^*)^2.$$

Notice that

$$\begin{aligned}(yy^*)\textstyle\sum[dy^i dy^{*i}] &- [(y^*dy)(ydy^*)] \\ &= \{(yy^*)(y'y^{*\prime}) - (y^*y')(yy^{*\prime})\}[dtdt^*] = ((y\wedge y')(y^*\wedge y^{*\prime}))[dtdt^*],\end{aligned}$$

where the notation $a \wedge b$ denotes the bivector with the components

$$\alpha^{ij} = a^i b^j - a^j b^i.$$

Consequently we have

$$\Omega^1 = \sum_{i=0}^{n} [dx^i dx^{*i}] = |\, y\wedge y' \,|^2 / |\, y \,|^4 [dtdt^*] \tag{5.1}$$

and finally, according to the definition (4.4),

$$J_1(C) = 1/2\pi i \int_C |\, y\wedge y' \,|^2 / |\, y \,|^4 [dtdt^*]. \tag{5.2}$$

If C is an algebraic curve, then (5.2) gives the order of C (except the sign which depends upon the orientation assumed for C).

Let us now consider the variety C_r generated by the $(r-1)$-osculating linear spaces of C. In order to evaluate $\int_{C_r} \Omega^r$ we consider two consecutive $(r-1)$-osculating spaces, $L_{r-1}(t)$, defined by the points $y(t), y'(t), \cdots, y^{(r-1)}(t)$, and $L_{r-1}(t+dt)$, defined by $y(t+dt), y'(t+dt), \cdots, y^{(r-1)}(t+dt)$, which are contained in the linear space $L_r(t)$ defined by the points $y(t), y'(t), \cdots, y^{(r)}(t)$. The angle $d\tau$ between $L_{r-1}(t)$ and $L_{r-1}(t+dt)$ is equal to the distance Ω^1 between the poles of $L_{r-1}(t)$ and $L_{r-1}(t+dt)$ considered as linear subspaces of L_r. Choosing the coordinate system so that the equations of L_r are $y^{r+1}=y^{r+2}=\cdots=y^n=0$, the pole of $L_{r-1}(t)$ in $L_r(t)$ will be the point Y_r whose coordinates are the determinants of order r in the r by $r+1$ matrix $(y^{k(i)})$, where $i=0,1,\cdots,r-1$, and $k=0,1,\cdots,r$.

We now apply (5.1). The coordinates of the point Y'_r are the determinants of order r of the r by $r+1$ matrix $(y^{k(i)})$, where $i=0,1,\cdots,r-2,r$, and $k=0,1,\cdots,r$, and following a device due to Ahlfors and H. Weyl ([1], [9], p. 144), one can prove that

$$|Y_r \wedge Y'_r| = |Y_{r-1}| \cdot |Y_{r+1}|, \tag{5.2}$$

where Y_{r-1} and Y_{r+1} have a meaning analogous to that of Y_r, i. e. they are the multivectors defined by $(y, y', \cdots, y^{(r-2)})$ and $(y, y', \cdots, y^{(r)})$ respectively.

The device consists of choosing a coordinate system in which $y, y', \cdots, y^{(r)}$ have the coordinates

$$y(y^0, 0, \cdots, 0), y'(y^{0'}, y^{1'}, 0, \cdots, 0), \cdots, y^{(r)}(y^{0(r)}, y^{1(r)}, \cdots, y^{r(r)}).$$

In this system of coordinates the relation (5.2) becomes trivial.

Therefore we have

$$d\tau = |Y_{r-1}|^2 \cdot |Y_{r+1}|^2 / |Y_r|^4 [dt dt^*].$$

Since according to (4.3) $\int d\tau = \int \Omega^1 = 2\pi i$ we have the relation

$$d\tau / 2\pi i = \Omega^r / \int_{L_r} \Omega^r$$

and consequently we get

$$J_r(C_r) = r!/(2\pi i)^r \int_{C_r} \Omega^r = 1/2\pi i \int_C d\tau \tag{5.3}$$

$$= 1/2\pi i \int_C |Y_{r-1}|^2 \cdot |Y_{r+1}|^2 / |Y_r|^4 [dt dt^*].$$

This expression is due to Ahlfors and H. Weyl [1], [9]. If C is an algebraic curve (5.3) give the classes of different orders.

For instance, for a plane algebraic curve $y^0 = y^0(t)$, $y^1 = y^1(t)$, $y^2 = y^2(t)$, the class is given (except for the sign) by the integral

$$J_2(C) = 1/2\pi i \int_C | y |^2 \cdot | yy'y'' |^2 / | y \wedge y' |^4 [dtdt^*],$$

where $| yy'y'' |$ denotes the absolute value of the determinant formed by the components of y, y', y''.

6. Total volume of the unitary group and of the homogeneous spaces U/Γ_r. In order to calculate the volume $\int_U du$ of the unitary group U with the invariant element of volume du given by (3.4), we put $\xi_0{}^k = \rho_k e^{i\theta_k}$ and consequently

$$\omega_{00} = (\xi^*{}_0 d\xi_0) = \sum_{k=0}^{n} (\rho_k d\rho_k + i\rho_k{}^2 d\theta_k).$$

Since $(\xi^*{}_0\xi_0) = \sum_0^n \rho_k{}^2 = 1$, we have $\sum_0^n \rho_k d\rho_k = 0$, and therefore the integral of ω_{00} over all possible values of the variables has the value

$$\int \omega_{00} = i \sum_{k=0}^{n} \int \rho_k{}^2 d\theta_k = 2\pi i.$$

The same proof gives

$$\int \omega_{jj} = 2\pi i. \tag{6.1}$$

On the other hand we observe that, according to (4.9), the exterior product

$$[\prod_{h=1}^{n-r} \omega_{r,r+h}\omega^*{}_{r,r+h}] = [\prod_{h=1}^{n-r} (\xi^*{}_{r+h} d\xi_r)(\xi_{r+h} d\xi^*{}_r)]$$

denotes the element Ω^{n-r} relative to the linear $(n-r)$-dimensional space defined by the points $\xi_r, \xi_{r+1}, \cdots, \xi_n$. Consequently (4.3) gives

$$\int [\prod_{h=1}^{n-r} \omega_{r,r+h}\omega^*{}_{r,r+h}] = (2\pi i)^{n-r}/(n-r)!. \tag{6.2}$$

From (3.4), (6.1) and (6.2) we get

$$\int_U du = (2\pi i)^{n+1} \prod_{r=0}^{n} (2\pi i)^{n-r}/(n-r)!$$

which can be written

$$\int_U du = \prod_{h=1}^{n+1} (2\pi i)^h/(h-1)!. \tag{6.3}$$

This formula gives the total volume of the unitary group U.

We wish now to calculate the total volume of the homogeneous space U/Γ_r, or, what is the same, the total measure of all the linear r-dimensional spaces contained in the hermitian n-dimensional space.

Let us write, for the moment, $du = du_n$, in order to exhibit clearly the dimension n of the space. According to (3.2) and (3.4), we have the following relation between the elements of volume du_n, dL_r, and the elements du_r, du_{n-r}, of the r- and $(n-r)$-dimensional unitary groups:

$$du_n = [du_r \, du_{n-r-1} \, dL_r].$$

Integration of both sides of this equality over all possible values of the variables, taking (6.3) into account, gives

$$\prod_{h=1}^{n+1} (2\pi i)^h/(h-1)! = \prod_{h=1}^{r+1} (2\pi i)^h/(h-1)! \prod_{h=1}^{n-r} (2\pi i)^h/(h-1)! \int_{U/\Gamma_r} dL_r,$$

and consequently

$$\int_{U/\Gamma_r} dL_r = (2\pi i)^{(n-r)(r+1)} 1!2! \cdots r!/[n!(n-1)! \cdots (n-r)!]. \tag{6.4}$$

This is the measure of all the linear r-dimensional spaces of the hermitian n-dimensional space, i.e. the total volume of the homogeneous space U/Γ_r.

The finite values (6.3) and (6.4) induce a normalization of the elements of volume du and dL_r in such a way that the total volumes of the corresponding spaces are equal to 1. These normalized elements will be

$$du' = \prod_{h=1}^{n+1} (h-1)!/(2\pi i)^h du \tag{6.5}$$

$$dL_r' = (2\pi i)^{-(n-r)(r+1)} n!(n-1)! \cdots (n-r)!/1!2! \cdots r! [dL_r]. \tag{6.6}$$

7. Linear subspaces which intersect an analytic variety. Let C_h be a fixed analytic variety of complex dimension h. Let L_r^0 be a fixed r-dimensional linear subspace and $L_r = uL_r^0$ be the transform of L_r^0 by $u \in U$. We assume $r + h - n \geq 0$, and let $C_h \cap L_r$ be the $(r+h-n)$-dimensional variety intersection of C_h with L_r. We wish to evaluate the integral

$$I = \int_{U/\Gamma_r} J_{h+r-n}(C_h \cap L_r) \, dL_r'. \tag{7.1}$$

Let ξ_0 be a point of $C_h \cap L_r$. In order to define L_r, we may take the points $\xi_0, \xi_1, \cdots, \xi_{h+r-n}$ on the linear $(h+r-n)$-space tangent to the intersection $C_h \cap L_r$ at ξ_0, and the points $\alpha_{h+r-n+1}, \cdots, \alpha_r$ such that

$$(7.2) \qquad (\xi_i \xi^*_j) = \delta_{ij}, \qquad (\xi^*_i \alpha_l) = 0, \qquad (\alpha^*_l \alpha_m) = \delta_{lm}.$$

Let $\gamma_{r+1}, \cdots, \gamma_n$ be $n-r$ points such that

$$(7.3) \qquad (\gamma^*_p \xi_i) = 0, \qquad (\gamma^*_p \alpha_l) = 0, \qquad (\gamma^*_p \gamma_q) = \delta_{pq},$$

and let $\beta_{h+r-n+1}, \cdots, \beta_h$ be $n-r$ points on the tangent space to C_h at ξ_0 such that

$$(7.4) \qquad (\beta^*_a \beta_b) = \delta_{ab}, \qquad (\beta^*_a \xi_i) = 0.$$

In (7.2), (7.3), and (7.4) we agree on the range of indices

$$0 \leq i, j \leq h+r-n, \quad h+r-n+1 \leq l, m \leq r, \quad r+1 \leq p, q \leq n$$
$$h+r-n+1 \leq a, b \leq h.$$

According to (3.2) and (2.2) we have

$$(7.5) \qquad dL_r = [\prod (\gamma^*_p d\xi_i)(\gamma_p d\xi^*_i) \prod (\gamma^*_p d\alpha_l)(\gamma_p d\alpha^*_l)],$$

the product being extended over all γ, ξ, α. Since $d\xi_0$ is on the tangent h-space to C_h, we have

$$d\xi_0 = \sum_i A_i \xi_i + \sum_a B_a \beta_a,$$

where, according to (7.2) and (7.4), $A_i = (\xi^*_i d\xi_0)$, $B_a = (\beta^*_a d\xi_0)$.

Consequently

$$(\gamma^*_p d\xi_0) = \sum_a (\gamma^*_p \beta_a)(\beta^*_a d\xi_0), \qquad (\gamma_p d\xi^*_0) = \sum_a (\gamma_p \beta^*_a)(\beta_a d\xi^*_0),$$

and

$$(7.6) \quad [\prod_p (\gamma^*_p d\xi_0)(\gamma_p d\xi^*_0)] = \| (\gamma^*_p \beta_a) \| \cdot \| (\gamma_p \beta^*_a) \| [\prod_a (\beta^*_a d\xi_0)(\beta_a d\xi^*_0)].$$

According to (4.9), the differential invariant Ω^h referred to C_h and the differential invariant Ω^{r+h-n} referred to $C_h \cap L_r$ may be written

$$(7.7) \qquad \Omega^h = [\prod_i (\xi^*_i d\xi_0)(\xi_i d\xi^*_0) \prod_a (\beta^*_a d\xi_0)(\beta_a d\xi^*_0)],$$

$$(7.8) \qquad \Omega^{h+r-n} = [\prod_i (\xi^*_i d\xi_0)(\xi_i d\xi^*_0)].$$

Consequently, from (7.5) and (7.6) we have

$$(7.9) \qquad [\Omega^{h+r-n} dL_r] = [\theta \Omega^h],$$

where θ is a differential form which does not depend upon ξ_0. According to (6.6) the same formula holds, up to a constant factor, for dL_r' instead of dL_r. The integral of θ over all possible values of the variables gives a constant value c which does not depend upon C_h. Consequently, integration of (7.9) gives $I = cJ_h(C_h)$. The value of c can be found by considering the case in which C_h is a linear subspace L_h. In this case we have $J_{h+r-n}(L_h \cap L_r) = J_{h+r-n}(L_{h+r-n}) = 1$, $J_h(L_h) = 1$, and $\int dL_r' = 1$. Consequently $c = 1$.

We get the final result

$$\int_{U/\Gamma_r} J_{r+h-n}(C_h \cap L_r)\, dL_r' = J_h(C_h). \tag{7.10}$$

If $r + h - n = 0$, $J_0(C_h \cap L_r)$ denotes the number of intersection points of L_r with C_h.

8. Analytic varieties which intersect each other. Let C_h, C_r $(h + r - n \geqq 0)$ be two analytic varieties of dimension h, r respectively. Let uC_r be the transform of C_r by $u \varepsilon U$. We consider the integral

$$I = \int_U J_{r+h-n}(C_h \cap uC_r)\, du'. \tag{8.1}$$

The frame which determines u may be chosen in the following way. Let ξ_0 be a point of the intersection $C_h \cap uC_r$. We choose the points $\xi_1, \xi_2, \cdots, \xi_{h+r-n}$ on the linear $(h + r - n)$-dimensional space tangent to the variety $C_h \cap uC_r$ at ξ_0, and the points $\alpha_{h+r-n+1}, \cdots, \alpha_r$ on the tangent r-space to C_r at ξ_0, in such a way that

$$(\xi^*_i\xi_j) = \delta_{ij}, \qquad (\xi^*_i\alpha_l) = 0, \qquad (\alpha^*_l\alpha_m) = \delta_{lm}. \tag{8.2}$$

Let $\gamma_{r+1}, \cdots, \gamma_n$ be $n - r$ points such that

$$(\gamma^*_p\xi_i) = 0, \qquad (\gamma^*_p\alpha_l) = 0, \qquad (\gamma^*_p\gamma_q) = \delta_{pq}. \tag{8.3}$$

In (8.2), (8.3), and in the remainder of this section we agree on the ranges of indices $0 \leq i, j \leq h + r - n$, $h + r - n + 1 \leq l, m \leq r$, $r + 1 \leq p, q \leq n$.

The frame which determines u is the n-simplex ξ_i, α_l, γ_p. According to (3.4) we have

$$\begin{aligned} du = [&\textstyle\prod(\xi^*_i d\xi_i)\prod(\alpha^*_l d\alpha_l)\prod(\gamma^*_p d\gamma_p)\prod(\xi^*_i d\xi_j)(\xi_i d\xi^*_j)\prod(\alpha^*_l d\alpha_m)(\alpha_l d\alpha^*_m) \\ &\textstyle\prod(\alpha^*_l d\xi_i)(\alpha_l d\xi^*_i)\prod(\gamma^*_p d\gamma_q)(\gamma_p d\gamma^*_q)\prod(\gamma^*_p d\xi_i)(\gamma_p d\xi^*_i)\prod(\gamma^*_p d\alpha_l)(\gamma_p d\alpha^*_l)], \end{aligned} \tag{8.4}$$

where $i \neq j$, $l \neq m$, $p \neq q$.

Let $\beta_{h+r-n+1},\cdots,\beta_h$ be $n-r$ points on the tangent h-space to C_h at ξ_0, such that

$$(\beta^*{}_a\beta_b)=\delta_{ab},\quad (\beta^*{}_a\xi_i)=0, \tag{8.5}$$

where $h+r-n+1\leq a,b\leq h$.

Since we always take ξ_0 on C_h, we have

$$d\xi_0=\sum_i A_i\xi_i+\sum_a B_a\beta_a$$

where, according to (8.2) and (8.3), $A_i=(\xi^*{}_i d\xi_0)$, $B_a=(\beta^*{}_a d\xi_0)$.

Consequently we have

$$(\gamma^*{}_p d\xi_0)=\sum_a(\gamma^*{}_p\beta_a)(\beta^*{}_a d\xi_0),\qquad (\gamma_p d\xi^*{}_0)=\sum_a(\gamma_p\beta^*{}_a)(\beta_a d\xi^*{}_0),$$

and by exterior multiplication,

$$[\prod_p(\gamma^*{}_p d\xi_0)(\gamma_p d\xi^*{}_0)]=\|(\gamma^*{}_p\beta_a)\|\cdot\|(\gamma_p\beta^*{}_a)\|[\prod_a(\beta^*{}_a d\xi_0)(\beta_a d\xi^*{}_0)]. \tag{8.6}$$

The differential invariant Ω^r referred to C_r, according to (4.9), is

$$\Omega^r=[\prod_i(\xi^*{}_i d\xi_0)(\xi_i d\xi^*{}_0)\prod_l(\alpha^*{}_l d\xi_0)(\alpha_l d\xi^*{}_0)]. \tag{8.7}$$

From (7.7), (7.8), (8.4), (8.6), and (8.7) we deduce

$$[\Omega^{h+r-n}du]=[\psi\Omega^r\Omega^h], \tag{8.8}$$

where ψ is a differential form which does not depend upon ξ_0. Up to a constant factor, according to (6.5), the same formula holds for du' instead of du. The integral of ψ over all possible values of the variables gives a constant value c. Consequently from (8.1) and (8.8) we deduce $I=cJ_h(C_h)J_r(C_r)$.

In order to determine the value of the constant c we consider the case in which C_h and C_r are linear spaces L_h and L_r. In this case we have $J_{h+r-n}(L_h\cap uL_r)=1$, $\int du'=1$, $J_h(C_h)=J_r(C_r)=1$. Consequently we have proved the integral formula

$$\int_U J_{r+h-n}(C_h\cap uC_r)\,du'=J_h(C_h)J_r(C_r). \tag{8.9}$$

If $r+h-n=0$, $J_0(C_h\cap uC_r)$ denotes the number of intersection points of C_h with uC_r.

For algebraic varieties, since the invariant J coincides with the order of the variety, (8.9) is an integrated form of the theorem of Bezout. Therefore (8.9) may be considered as a generalization of the theorem of Bezout to analytic varieties.

If we have $p+1$ analytic varieties $C_{h_0}, C_{h_1}, \cdots, C_{h_p}$ such that $h_0+h_1+\cdots+h_p \geq np$, from (8.9) we get immediately by recurrence

$$\int_U J_{h_0+h_1+\cdots+h_p-np}(C_h \cap u_1 C_{h_1} \cap u_2 C_{h_2} \cap \cdots \cap u_p C_{h_p})\, du_1' du_2' \cdots du_p'$$
$$= J_{h_0}(C_{h_0}) J_{h_1}(C_{h_1}) \cdots J_{h_p}(C_{h_p}).$$

THE UNIVERSITY OF LA PLATA, ARGENTINA.

BIBLIOGRAPHY.

[1] L. V. Ahlfors, "The theory of meromorphic curves," *Acta Societatis Scientiarum Fennicae*, A, vol. III, 1941.

[2] W. Blaschke, "Densita negli spazi di Hermite," *Rendiconti dell' Academia dei Lincei*, vol. 29, 6, pp. 105-108, 1939.

[3] E. Cartan, *La théorie des groupes finis et continus et la géométrie differentielle traités par la méthode du repere mobile*, Paris, 1937.

[4] E. Cartan, "Sur les invariants intégraux de certains espaces homogènes clos et les proprietés topologiques de ces espaces," *Annales de la Société Polonaise des Mathématiques*, vol. 8, pp. 181-225, 1929.

[5] S. S. Chern, "On integral geometry in Klein spaces," *Annals of Mathematics*, vol. 43, pp. 178-189, 1942.

[6] H. Rohde, "Unitäre Integralgeometrie," *Hamburger Abhandlungen*, vol. 13, pp. 295-318, 1940.

[7] L. A. Santaló, "Integral Geometry in projective and affine spaces," *Annals of Mathematics*, vol. 51, pp. 739-755, 1950.

[8] O. Varga, "Ueber die Integralinvarianten die zu einer Kurve in der Hermitischen Geometrie gehören," *Acta Litt. Scientarum Szeged*, vol. 9, pp. 88-102, 1939.

[9] H. Weyl, "Meromorphic functions and analytic curves," *Annals of Mathematics Studies*, No. 12, Princeton, 1943.

ON THE PARABOLIC GENERATORS OF THE PRINCIPAL CONGRUENCE SUBGROUPS OF THE MODULAR GROUP.*

By Emil Grosswald.

1. The set of matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, where the integers a, b, c, d satisfy

$$ad - bc = 1 \tag{1}$$

and where we consider $\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} -a & -b \\ -c & -d \end{pmatrix}$, form the modular group. We define $\Gamma(p)$, the principal congruence subgroup of Γ modulo a prime p, by the additional conditions $a \equiv d \equiv 1$ and $b \equiv c \equiv 0$ (unless otherwise stated, all congruences are taken $\bmod p$). For $p > 3$, H. Frasch has shown ([3]; see also [7]) that $\Gamma(p)$ can be generated as a free group by the generators $S^p = \begin{pmatrix} 1 & p \\ 0 & 1 \end{pmatrix}$ and $(p-1)p(p+1)/12$ other generators, depending on three parameters, identified by the symbols (λ, μ, ν) and defined as follows. Let

$$V = TS^{\alpha}TS^{\beta}TS^{\alpha} = \begin{pmatrix} \beta & \alpha\beta - 1 \\ -(\alpha\beta - 1) & -\alpha(\alpha\beta - 2) \end{pmatrix}, \tag{2}$$

where α and β are any two primitive roots $(\bmod p)$ satisfying $\alpha\beta \equiv 1$. Let λ, μ, ν be any set of integers satisfying $0 \leqq \lambda \leqq (p-3)/2$, $0 \leqq \mu \leqq p-1$, $1 \leqq \nu \leqq p-1$, and define

$$\begin{gathered} \sigma = \operatorname*{ind}_{\alpha} \nu, \qquad \lambda_* = \lambda + \sigma \quad (\text{reduced mod } \tfrac{1}{2}(p-1)), \\ \mu_* = \mu\nu^2 - \nu \quad (\text{reduced mod } p), \quad \nu\nu_* \equiv -1 \quad (\text{mod } p); \end{gathered} \tag{3}$$

so that λ_*, μ_*, ν_* are in the same range, respectively, as λ, μ, ν. Let us call "reciprocals" two numbers related as ν and ν_*. Finally, let

$$\begin{aligned} R &= S^{\mu}TS^{\nu}TS^{-\nu_*}TS^{-\mu_*} \\ &= \begin{pmatrix} -\mu(\nu\nu_* + 1) + \nu_* & \mu\mu_*(\nu\nu_* + 1) - (\mu\nu + \mu_*\nu_*) + 1 \\ -(\nu\nu_* + 1) & \mu_*(\nu\nu_* + 1) - \nu \end{pmatrix}. \end{aligned} \tag{4}$$

Then the generators (λ, μ, ν) of $\Gamma(p)$ are defined by

* Received January 12, 1951.

$$(\lambda, \mu, \nu) = V^{\lambda} R V^{-\lambda_*}. \tag{5}$$

The object of this paper is to show that if the prime p satisfies a certain condition [1] to be stated later, the independent generators (λ, μ, ν) can be chosen in such a way that none of them is parabolic,[2] so that the only parabolic generator of $\Gamma(p)$ is S^p.

For our proof, we have to change slightly the process of elimination indicated by Frasch [3], whose method we follow, however, closely. Throughout the paper we consider only values of $p > 7$ (cf. a similar situation in [7], section I); in the other cases we would need sometimes special considerations, which complicate needlessly the exposition, as the corresponding generators are explicitly indicated in [3] and none of them is parabolic. A few inequalities will be meaningful only for some higher values of p and will be understood to apply only for them.

2. We recall some results of Frasch [3], with some of their immediate consequences. The $\frac{1}{2}p(p-1)^2$ generators (λ, μ, ν) satisfy $p(p-1)(5p-7)/12$ distinct defining relations, each permitting the elimination of one generator. We use them in order to eliminate all generators with a same value ν_0 of ν, unless ν_0 satisfies

$$\nu_0^2 \equiv -1 \quad \text{or} \tag{6}$$

$$\nu_0(\nu_0 - 1) \equiv -1. \tag{6'}$$

The defininig relations can be arranged in groups of binary and ternary relations. Following Frasch, we use the first group of binary relations to eliminate the generators with $\nu = 1$. The following ternary relations permit the elimination of $\nu = p - 1$. Using the second group of binary relations, $\nu = 2$ and $\nu = p - 2$ are eliminated; and the corresponding ternary relations permit the elimination of their reciprocals, $\nu = \frac{1}{2}(p \pm 1)$. In general, for $k \geqq 3$, we use the binary group to eliminate $\nu = k$, $\nu = p - k$, and $\nu = \frac{1}{2}\{p \pm (2k-3)\}$, and the corresponding ternary relations to eliminate their reciprocals. The generators with the mentioned values of ν actually appear in the defining relations in the stated order, as explicitly shown in [3]. The number of relations at our disposal is larger than the number of generators on which we have centered our attention, and we may use the remaining ones in the order suggested in [3]. The process of elimination

---

[1] It seems likely that all primes satisfy that condition; see last footnote.

[2] A corresponding property holds for the generators V_k of $\Gamma_0(p)$, defined in [9]. See [5].

has to be modified for $p \not\equiv -1 \pmod{12}$, when ν takes on values satisfying (6) or (6′). Let $\nu = k$ or $\nu = p - k$ satisfy (6). Then $k^2 + 1 \equiv 0$, so that $k^2 - (p-1) \geqq 0$, and $k \geqq (p-1)^{\frac12}$. If ν satisfies (6′), then $k(k-1) \equiv -1$, so that $k^2 - k - (p-1) \geqq 0$, $k \geqq \frac12 + (p - \frac34)^{\frac12} > (p-1)^{\frac12}$. In the same way we show that if $\nu = \frac12\{p \pm (2k-3)\}$ satisfies (6) or (6′), then $k > \frac12(p^{\frac12} + 1)$. It follows that all values of ν belonging to any one of the three intervals

$$(7) \qquad 1 \leqq \nu \leqq p^{\frac12} - 1; \quad \tfrac12(p - p^{\frac12}) + 1 < \nu < \tfrac12(p + p^{\frac12}) - 1; \quad p - p^{\frac12} + 1 < \nu \leqq p - 1$$

are not affected by the difficulties arising from (6), (6′), and can be eliminated from the final set of independent generators, using only the binary relations. Whenever we reach any value satisfying (6), or (6′), we continue as in [3]. If ν does not satisfy (6) or (6′), respectively, then the same is true of its reciprocal, and the generators with values of ν reciprocal to those of (7) will also be eliminated, using the corresponding ternary relations. The values of ν in (7) and their reciprocals form a set of at most $3(p^{\frac12} - 1)$ distinct numbers. For the elimination of the corresponding generators we dispose (for some fixed λ and μ) of $\frac{1}{12}p(p-1)(5p-7)/\frac12 p(p-1) = (5p-7)/6$ relations. For $p > 7$, $(5p-7)/6 \geqq 3(p^{\frac12} - 1)$, proving the previous assertion that the number of independent defining relations exceeds the number of generators which we want to eliminate. Finally, we remark that there remain in the generators only values of ν, which satisfy, with their reciprocals, the inequalities

$$(8) \quad p^{\frac12} - 1 < \nu, \nu_* < \tfrac12(p - p^{\frac12}) + 1, \text{ or } \tfrac12(p + p^{\frac12}) - 1 < \nu, \nu_* < p - p^{\frac12} + 1.$$

3. From (3) and (5) it follows that

$$(9) \quad \begin{aligned} V^{-\lambda}(\lambda, \mu, \nu) V^{\lambda} &= RV^{-\sigma}, \quad \text{if} \quad \lambda + \sigma < \tfrac12(p-1), \quad \text{and} \\ V^{-\lambda}(\lambda, \mu, \nu) V^{\lambda} &= RV^{-\sigma + \frac12(p-1)}, \quad \text{if} \quad \lambda + \sigma \geqq \tfrac12(p-1). \end{aligned}$$

If (λ, μ, ν) is parabolic, then all its transforms $M^{-1}(\lambda, \mu, \nu)M$ by any modular matrix M are also parabolic. In order to prove that (λ, μ, ν) is not parabolic it is, therefore, sufficient to show that $RV^{-\sigma}$ and $RV^{-\sigma+\frac12(p-1)}$ are not parabolic. Let, generally, $V^n = \begin{pmatrix} A_n & B_n \\ C_n & D_n \end{pmatrix}$. Then it follows from $V^{n+1} = V^n \cdot V = \begin{pmatrix} A_{n+1} & B_{n+1} \\ C_{n+1} & D_{n+1} \end{pmatrix}$ that

$$(10)\qquad \begin{aligned} A_{n+1} &= \beta A_n - (\alpha\beta - 1)B_n \qquad & B_{n+1} &= (\alpha\beta - 1)A_n - \alpha(\alpha\beta - 2)B_n \\ C_{n+1} &= \beta C_n - (\alpha\beta - 1)D_n \qquad & D_{n+1} &= (\alpha\beta - 1)C_n - \alpha(\alpha\beta - 2)D_n. \end{aligned}$$

For $n = 1$, by (2), $A_1 - D_1 = B_1(\alpha - (\alpha - \beta)/(\alpha\beta - 1))$ and $B_1 + C_1 = 0$. Assuming that

$$(11)\qquad A_n - D_n = 2tB_n, \text{ where } t = \tfrac{1}{2}(\alpha - (\alpha - \beta)/(\alpha\beta - 1)),$$

$$(12)\qquad B_n + C_n = 0$$

hold for n, it follows by induction on n, using (10), that (11) and (12) hold for $n + 1$; they hold, therefore, for all n. It follows from (1) and (12) that

$$(13)\qquad A_nD_n + B_n^2 = 1$$

and, as B_n is an integer,

$$(14)\qquad A_nD_n \leqq 0.$$

Hence, by (11) and (12),

$$(15)\qquad A_nB_n > 0, \qquad A_nC_n < 0.$$

For $n = 1$, by (2), $B_1 > A_1$; assuming that, for arbitrary integral n,

$$(16)\qquad | B_n | > | A_n |,$$

it follows by induction on n, using (10) and (15), that (16) holds for $n + 1$ and so for all n. From (16) it obviously follows that $(\alpha\beta - 1)B_n/A_n - \beta > 0$ so that, by (10), $A_{n+1}A_n < 0$. Hence, using (15), (12) and (14), $B_nB_{n+1} < 0$, $C_nC_{n+1} < 0$ and $D_nD_{n+1} \leqq 0$. We shall see immediately that the equality cannot hold, so that we have, generally, with M standing for A, B, C or D,

$$(17)\qquad M_nM_{n+1} < 0.$$

Let, generally, $\alpha\beta = Kp + 1$, $K \geqq 1$; then $B_1 = Kp$. From (10) it follows, by (14) and (17), that $| B_{n+1}/B_n | = \alpha(\alpha\beta - 2) - (\alpha\beta - 1)A_n/B_n > \alpha(\alpha\beta - 2) - (\alpha\beta - 1) = \alpha(Kp - 1) - Kp \geqq Kp - 2$ so that $| B_n | > | B_1 | (Kp - 2)^{n-1} = K^n p^n (1 - 2/Kp)^{n-1}$ and, for $p > 7$,

$$(18)\qquad | B_n | > (5p/7)^n.$$

Hence it follows in particular that $|B_n| > 1$ and the equality signs in (14) and in $D_n D_{n+1} \leqq 0$ may be suppressed. We have, consequently, using (15) and (12),

$$A_n D_n < 0, \qquad B_n D_n < 0, \qquad C_n D_n > 0. \tag{19}$$

We solve now (11) and (12) for A_n and D_n, using (16) and (19) to determine the signs of the radicals and obtain

$$A_n = B_n(t - (t^2 - 1 + B_n^{-2})^{\frac{1}{2}})$$

$$D_n = -B_n(t + (t^2 - 1 + B_n^{-2})^{\frac{1}{2}}).$$

As, by (11) and (18), $2 < 2t < B_n$, it follows that

$$t - 1/2t > (t^2 - 1 + B_n^{-2})^{\frac{1}{2}} > t - 1/(2t-1),$$

so that

$$\begin{gathered} 2t - 1/(2t-1) < |D_n/B_n| < 2t - 1/2t \\ 1/2t < A_n/B_n < 1/(2t-1). \end{gathered} \tag{20}$$

4. Let now, for integral $n \geqq 1$, $RV^{-n} = R\begin{pmatrix} D_n & -B_n \\ -C_n & A_n \end{pmatrix} = \begin{pmatrix} \mathfrak{A}_n & \mathfrak{B}_n \\ \mathfrak{C}_n & \mathfrak{D}_n \end{pmatrix}$.

If this matrix is parabolic, then we have (see [2, p. 21]) $|\mathfrak{A}_n + \mathfrak{D}_n| = 2$, where, by (4),

$$\begin{aligned} \mathfrak{A}_n &= -D_n[\mu(\nu\nu_* + 1) - \nu_*] - C_n[\mu\mu_*(\nu\nu_* + 1) - (\mu\nu + \mu_*\nu_*) + 1] \\ \mathfrak{D}_n &= B_n(\nu\nu_* + 1) + A_n[\mu_*(\nu\nu_* + 1) - \nu]. \end{aligned} \tag{21}$$

If $\mu\mu_* \neq 0$, all brackets are positive, as is obvious for all except the coefficient of C_n. Let its value be N, and apply (1) to R. It follows that $N > 0$, and by (15), (18) and (19), we obtain $|\mathfrak{A}_n + \mathfrak{D}_n| > |B_n|(N + \nu\nu_* + 1) > (5p/7)^n(N + p) > 2$, so that RV^{-n} cannot be parabolic. If $\mu = 0$, then, by (3), $\mu_* = p - \nu$. From (15) and (19) it follows that $|\mathfrak{A}_n + \mathfrak{D}_n| = ||\mathfrak{A}_n| - |\mathfrak{D}_n||$; and by (20) and (21), we have $l < |\mathfrak{A}_n/B_n| < L$, $H > |\mathfrak{D}_n/B_n| > h$, with

$$l = (p - \nu)\nu_* - 1 + [2t - 1/(2t-1)]\nu_*,$$
$$L = (p - \nu)\nu_* - 1 + (2t - 1/2t)\nu_*,$$

$h = (\nu\nu_* + 1) + [(p - \nu)(\nu\nu_* + 1) - \nu]/2t,$

$$H = (\nu\nu_* + 1) + [(p - \nu)(\nu\nu_* + 1) - \nu]/(2t - 1).$$

Consequently,

$$(22) \qquad |\mathfrak{D}_n| - |\mathfrak{A}_n| > |B_n|(h - L)$$

and we assert that

$$(23) \qquad h - L > 1/2t$$

if

$$(24) \qquad \text{either } \beta < \alpha \leqq \nu \text{ or } \alpha \leqq \nu - 1, \beta > \alpha.$$

Proof. (23) is equivalent to

$$4t^2 - 2t[2(\nu + 1/\nu_*) - p] - [(p - \nu)(\nu + 1/\nu_*) - (\nu + 1)/\nu_* + 1] \leqq 0,$$

which is satisfied for $0 \leqq t \leqq \frac{1}{2}\nu$, or, by (11), for $\alpha \leqq \nu + (\alpha - \beta)/(\alpha\beta - 1)$, proving the assertion.

By virtue of (8), (24) is always satisfied for

$$(25) \qquad \alpha \leqq p^{\frac{1}{2}} - 2$$

and also for

$$(25') \qquad \alpha \leqq \tfrac{1}{2}(p + p^{\frac{1}{2}}) - 2, \text{ if } \tfrac{1}{2}(p + p^{\frac{1}{2}}) - 1 < \nu < p - p^{\frac{1}{2}} + 1.$$

Similarly,

$$(26) \qquad |\mathfrak{A}_n| - |\mathfrak{D}_n| > |B_n|(l - H)$$

where

$$(27) \qquad l - H \geqq 1/(2t - 1) \text{ if } \qquad (28) \qquad \alpha \geqq \nu + 2.$$

Proof. (27) is equivalent to

$$4t^2 - 2t[2(\nu + 1/\nu_*) - p + 1]$$
$$- [(p - \nu - 1)(\nu + 1/\nu_*) - (\nu - 2)/\nu_* + p - \nu + 1] \geqq 0,$$

which is satisfied for $2t \geqq \nu + 1/\nu_* + 1$, or, by (11), for $\alpha \geqq \nu + 1/\nu_* + 1 - (\alpha - \beta)/(\alpha\beta - 1)$ proving the assertion. In view of (8), α satisfies (28), provided that

$$(29) \qquad \alpha \geqq p - p^{\frac{1}{2}} + 3$$

and also for

$$(29') \qquad \alpha \geqq \tfrac{1}{2}(p - p^{\frac{1}{2}}) + 3, \text{ if } p^{\frac{1}{2}} - 1 < \nu < \tfrac{1}{2}(p - p^{\frac{1}{2}}) + 1.$$

Combining (25′) and (29′), it follows that for any possible value of ν either (23) or (27) holds, in case we have simultaneously [3]

$$(30) \qquad \tfrac{1}{2}(p - p^{\frac{1}{2}}) + 3 < \alpha < \tfrac{1}{2}(p + p^{\frac{1}{2}}) - 2.$$

We shall say that a prime p has the property C, in case [4] at least one of the following three alternatives holds:

(a) the smallest primitive root of p does not exceed $p^{\frac{1}{2}} - 2$;

(b) the largest primitive root of p is not less than $p - p^{\frac{1}{2}} + 3$;

(c) there exists a primitive root of p in the interval (30).

Let us choose for α a primitive root satisfying either (a), or (b), or (c). From (22), (23), (26), and (27) it follows now, on account of (18), that if p has the property C, then

$$(31) \qquad | \mathfrak{A}_n + \mathfrak{D}_n | > | B_n |/2t > (5p/7)^n/p.$$

By (3) $\alpha^\sigma \equiv \nu$ and, by the choice of α, $\nu \neq \alpha$; consequently, $\sigma \neq 1$ in (9), and it is sufficient to consider in (31) only values $n \geqq 2$. As for $n \geqq 2$, $(5p/7)^n/p > 2$, it follows from (31) that $| \mathfrak{A}_n + \mathfrak{D}_n | > 2$, proving that if $\mu = 0$, then RV^{-n} is not parabolic. The case $\mu_* = 0$ can be reduced to the previous one by interchanging ν and ν_*, which is possible without any further change of the argument or of the chosen value of α, since ν and ν_* were both eliminated from the intervals (7). This finishes the proof that RV^{-n} is not parabolic.

Similarly, let $RV^n = \begin{pmatrix} \mathfrak{A}'_n & \mathfrak{B}'_n \\ \mathfrak{C}'_n & \mathfrak{D}'_n \end{pmatrix}$. Assuming that this matrix is parabolic, we obtain, proceeding as before,

---

[3] (25), (29) and (30) may be improved to $\alpha < p^{\frac{1}{2}}$, $p - \alpha \leqq p^{\frac{1}{2}}$ and $| \alpha - \frac{1}{2}p | \leqq \frac{1}{2}p^{\frac{1}{2}}$ respectively, by sharper estimates in (20). We remark also that (30) is meaningful only for $p \geqq 29$. For all smaller primes, however, either (25), or (29) hold. See also remark at the end of 1.

[4] Although it seems likely, that every prime has the property C, the best that is known at present is a weaker result of Loo-keng Hua [6]; see also [1], [4], and [8]. In [7] J. Nielsen considers only primes p, admitting as primitive roots either 2, or $p - 2$ (consequently also $\frac{1}{2}(p \pm 1)$). Property C represents a weakened Nielsen condition and every prime, satisfying Nielsen's condition, obviously has the property C.

$$
(21') \quad \begin{aligned} |\,\mathfrak{A}'_n + \mathfrak{D}'_n\,| = -A_n[\mu(\nu\nu_* + 1) - \nu_*] \\ + C_n[\mu\mu_*(\nu\nu_* + 1) - (\mu\nu + \mu_*\nu_*) + 1] \\ - B_n(\nu\nu_* + 1) + D_n[\mu_*(\nu\nu_* + 1) - \nu] = 2. \end{aligned}
$$

We observe that, except for its sign, (21′) is obtained from the values (21), interchanging the couples (μ, ν) and (μ_*, ν_*). Consequently, $|\,\mathfrak{A}_n + \mathfrak{D}_n\,| \neq 2$ implies $|\,\mathfrak{A}'_n + \mathfrak{D}'_n\,| \neq 2$ contrary to (21′), and RV^n is not parabolic, unless $n = 1$. From $\frac{1}{2}(p+1) - \sigma = 1$ would follow, however, by (3), $\sigma = 0$ and, further, $\nu = \alpha^\circ = 1$, contradicting (7). Therefore, in (9), the exponent of V is different from one and $RV^{\frac{1}{2}(p+1)-\sigma}$ is not parabolic. By (9) we conclude that no generator (λ, μ, ν) is parabolic, so that the only parabolic generator of $\Gamma(p)$ is S^p, q. e. d.

Remark. A separate consideration of the cases $p \equiv \pm 1, \pm 5 \pmod{12}$ yields some additional information and suggests that the above result holds for all primes, whether they have property C or not, with at most a finite number of exceptions. As an example, let $p \equiv -1 \pmod{12}$. Then no relations (6), (6′) exist, and all values of ν of the form $\nu = k$, $\nu = p - k$, $1 \leqq k \leqq [(5p-7)/24]$, and their reciprocals can be eliminated from the generators. Erdös has shown [1] that every prime p has a primitive root $\alpha < (\log p)^{17} p^{\frac{1}{2}}$; and the considerations of **4** hold, provided $\alpha \leqq \nu - 1$, or

$$
(32) \qquad (\log p)^{17} p^{\frac{1}{2}} \leqq [(5p-7)/24] - 1,
$$

which is true for all sufficiently large p. In the same way, results of Fueter [4] could be used to improve our results in the cases $p \equiv +1$, and $p \equiv \pm 5 \pmod{12}$. The interest of such considerations is, however, considerably diminshed by the fact that "sufficiently large" means, for (32), $p > e^{200}$. Therefore, the additional results obtained do not seem to justify a detailed case distinction (mod 12). It is rather to be expected that the known results concerning the smallest primitive root (mod p) will be improved sufficiently (see [8]) to show that all primes have the property C.

UNIVERSITY OF SASKATCHEWAN.

BIBLIOGRAPHY.

[1] P. Erdős, "On the least primitive root of a prime," *Bulletin of the American Mathematical Society*, vol. 51 (1945), pp. 131-132.

[2] L. R. Ford, *Automorphic Functions*, New York, 1929.

[3] H. Frasch, "Die Erzeugenden der Hauptkongruenzgruppen für Primzahlstufen," *Mathematische Annalen*, vol. 108 (1933), pp. 229-252.

[4] R. Fueter, "Ueber primitive Wurzeln von Primzahlen," *Commentarii Mathematicae Helv.*, vol. 18 (1946), pp. 217-223.

[5] E. Grosswald, "On the genus of the fundamental region of some subgroups of the modular group," *American Journal of Mathematics*, vol. 74 (1952), pp. 86-88.

[6] L. K. Hua, "On the least primitive root of a prime," *Bulletin of the American Mathematical Society*, vol. 48 (1942), pp. 726-730.

[7] J. Nielsen, "A study concerning the congruence subgroups of the modular group," *Det Kgl. Danske Videnskabernes Selskab. Matematisk-Fysiske Meddelelsen*, vol. 25, no. 18 (1950).

[8] S. Pillai, "On the smallest primitive root of a prime," *Journal of the Indian Mathematical Society* (New Series), vol. 8 (1944), pp. 14-17.

[9] H. Rademacher, "Ueber die Erzeugenden von Kongruenzuntergruppen der Modulgruppe," *Abhandlungen aus dem Seminar der Hamburgischen Universität*, vol. 7 (1929), pp. 134-138.

ON SOLUTIONS OF LINEAR PARTIAL DIFFERENTIAL EQUATIONS OF MIXED TYPE.*

By Stefan Bergman.

1. Integral operators in the theory of linear partial differential equations. The realization that a number of relations between some properties of solutions of a homogeneous linear partial differential equation [1]

$$L(\psi) = \psi_{z\bar{z}} + A\psi_z + B\psi_{\bar{z}} + F\psi = 0, \tag{1.1}$$

$z = x + iy,\ \bar{z} = x - iy,$

$$(\partial\psi/\partial z) = (1/2)[(\partial\psi/\partial x) - (\partial\psi/\partial y)],$$
$$(\partial\psi/\partial \bar{z}) = (1/2)[(\partial\psi/\partial x) + i(\partial\psi/\partial y)],$$

in the case where A, B, F are entire functions of their arguments, are either independent of the coefficients A, B, F of the equation $\boldsymbol{L}$, or depend only upon some properties of A, B, F, has been one of the most significant achievements of the theory of integral operators of the first kind. See [3, 5, 6].[2]

By the operation

$$\psi = \boldsymbol{O}(f) = \mathrm{Re}[\mathrm{P}(f; z_0)]; \tag{1.2}$$

$$\mathrm{P}(f; z_0) = \int_{t=-1}^{+1} E(z', \bar{z}', t) f(z'(1-t^2))\, dt/(1-t^2)^{1/2}$$

$$z' = z - z_0,\ \bar{z}' = \bar{z} - \bar{z}_0, \qquad \mathrm{Re} = \text{Real part},$$

where $E(z', \bar{z}', t)$ (the generating function of the integral operator) is a fixed function, an arbitrary function f of one complex variable is transformed into a solution ψ of (1.1).

Remark. The point (z_0), which appears as an arbitrary parameter in the

---

* Received March 7, 1951. Research paper done under Navy Contract Nord 10-449, Task 3.

The ideas expressed in this paper represent the personal views of the author, and are not necessarily those of the Bureau of Ordnance.

[1] For our purposes, it is convenient to use complex notation and sometimes to continue x, y to complex values of the arguments, even if our final results refer to the behavior of the solutions in the real domain, i. e., for real values of x and y.

[2] The numbers in brackets refer to the bibliography. A familiarity with previous publications except some portions of [9] is not assumed in the present paper.

operator, will be called its *point of reference*. Usually we set $z_0 = \bar{z}_0 = 0$, unless, as in the case in section 5, it is of essential importance to have $z_0, \bar{z}_0$ arbitrary. The possibility of using an arbitrary point (z_0) as a reference point becomes important if we consider functions f which are singular at the origin, or if the coefficients of the equation become singular there. (This is the case, for instance, for the equations considered subsequently).

To every equation (1.1), it is possible to find a function $E = E_{\mathrm{I}}$ (the *generating function of the first kind*) which is determined in a unique way, so that the operator $\boldsymbol{O}_1 = \mathrm{Re}[\mathrm{P}_1]$ preserves many properties of the operator "Re."

Using this approach, it has been possible to show that various chapters of the theory of functions of a complex variable can be interpreted not only as theorems on harmonic functions, but also as theorems on solutions of equation (1.1) with entire coefficients A, B, F.

In particular [8] it has been shown that relations between the properties of the subsequence $\{A_{n0}\}$, $n = 0, 1, 2, \cdots$, of the coefficients of the series development $\psi = \sum_{m,n=0}^{\infty} A_{mn} z^m \bar{z}^n$ of a real solution ψ of $\boldsymbol{L}(\psi) = 0$, and the location and properties of its singularities are independent of the coefficients A, B, F, provided that these latter functions are entire in z and $\bar{z}$.

It is of interest that the method of transforming functions of one variable into solutions of linear differential equations can be extended to certain equations of mixed type. This is of particular importance for the theory of compressible fluids, which is based on the study of equations of this type. In this theory we investigate the equation

$$\boldsymbol{M}(\psi) \equiv \psi_{\mathrm{HH}} + l(\mathrm{H})\psi_{\theta\theta} = 0, \tag{1.3}$$

$$l(\mathrm{H}) = \sum_{n=1}^{\infty} a_n(-\mathrm{H})^n, \quad a_1 > 0, \quad a_2 < 0, \tag{1.4}$$

where $l(\mathrm{H}) > 0$ for $\mathrm{H} < 0$, and $l(\mathrm{H}) < 0$ for $\mathrm{H} > 0$. Reducing (1.3) to the normal form, one obtains the equation

$$\boldsymbol{L}(\psi) \equiv \psi_{z\bar{z}} + N(\lambda)(\psi_z + \psi_{\bar{z}}) = 0, \quad N = (1/8)l^{-3/2}l_{\mathrm{H}}, \quad \bar{Z} = \lambda - i\theta, \tag{1.1a}$$

$$-\lambda(-\mathrm{H}) = \int_{t=0}^{-\mathrm{H}} [l(t)]^{-1/2} dt, \tag{1.5}$$

[8] For other properties such as the existence of classes analogous to polynomials, approximation developments, summation theorems, the reader is referred to [3, 5, 6].

where $N(\lambda)$ is now singular[4] for $\lambda = 0$. Assuming that l and therefore N is an analytic function of the variable λ, it is convenient to continue our functions to complex values of the arguments. The domains $[-\infty < \theta < \infty, \mathrm{H} < 0]$ and $[-\infty < \theta < \infty, \mathrm{H} > 0]$ in the real plane correspond to the domains $[-\infty < \theta < \infty, \operatorname{Re} \lambda_1 < 0, \operatorname{Im} \lambda_1 = 0]$ and $[-\infty < \theta < \infty, \operatorname{Im} \lambda_1 > 0, \operatorname{Re} \lambda_1 = 0]$ respectively. Here $\lambda_1 = \lambda + i\Lambda$.

Since the integral operator of the first kind is not defined at the points where the coefficient N becomes singular, it is useful, in studying (1. 1a) to introduce two *integral operators* P_{21} and P_{22} (see (5. 8), (4. 1), and (4. 2)) *of the second kind*, which are defined in two different domains of the λ, θ-plane. Each of these operators is defined only in subdomains of the regularity domain of N, but P_{22} is defined also on the singularity line $\lambda = 0$ of N. P_{21} can be defined[5] for $\lambda \to -\infty$.

As is the case of regular coefficients A, B, F, it is possible to show that for a class of equations (1. 1a), namely those for which $N(\lambda)$ satisfies the conditions (3. 5a)-(3. 5d), a number of relations are either independent of $N(\lambda)$ or depend only on some properties of $N(\lambda)$. Among equations of this type, it is possible to select one, to which we shall refer as the "simplified case," where the situation is particularly simple, and whose theory can be used as a pattern for the development of the theory in the general case of equations of this type. In the "simplified case" where $l(\mathrm{H}) = -a_1\mathrm{H}$, a_1 a positive constant, the generating function[6] $E_2\dagger$ of the integral operator of the second kind is given by

$$(1.6) \qquad E_2\dagger = H\dagger E_2{}^*\dagger(Z, \bar{Z}, t), \quad H\dagger = S_0(-2\lambda)^{-1/6}, \quad Z = \lambda + i\theta,$$

where $E_2{}^*\dagger$ is a hypergeometric function of the variable $u = t^2Z/2\lambda$,

$$(1.7) \qquad E_2{}^*\dagger = A_1F(1/6, 5/6, 1/2, u) + B_1u^{1/2}F(2/3, 4/3, 3/2, u).$$

A_1 and B_1 are constants.

By the transformation $E_2{}^* = H^{-1}E_2$, equation (3. 2) of [9] for E_2 is reduced to the somewhat simpler equation (4. 9) for $E_2{}^*$.

---

[4] The study of the equations of mixed type is in this and many other respects equivalent to the study of partial differential equations of the form (1. 1) with singular coefficients.

[5] These results are of importance especially in fluid dynamics. To $\lambda = -\infty$ corresponds in the hodograph plane (i. e., in the plane whose Cartesian coordinates are the components u, v of the velocity) the origin $u = v = 0$.

[6] The generating function and some other quantities referring to the simplified case will be denoted by symobls with a dagger (†).

It has been shown in the general case that E_2^* satisfies a *partial differential equation* (see 4. 9), which in the "simplified case" becomes an *ordinary* differential equation (the hypergeometric equation). Equation (4. 9) possesses some features which can be considered a generalization of properties of hypergeometric equation.

In particular E_2^*† can be developed into a series of powers of u (and $v = 1 - u$), converging in $|Z/2\lambda| < 1$, as well as into a series of powers of u^{-1} (and $(1 - v^{-1})$) converging in $|Z/2\lambda| > 1$. Continuing the investigations of [9] in sections **4** and **6**, we show that in the general case of equation (1. 1a) there exist generating functions E_{21} and E_{22} representable in the form of infinite series, converging in $|Z/2\lambda| < 1$ and in $|Z/2\lambda| > 1$, respectively. They can be considered as generalizations of the above-mentioned series developments for the hypergeometric function. Substituting E_{21} and E_{22} into (1. 2) instead of E, we define the integral operators P_{21} and P_{22} in $|Z/2\lambda| < 1$ and in $|Z/2\lambda| > 1$, respectively.

Thus in the course of our investigations, we introduce a variety of integral operators which are defined in different domains, and which preserve different properties of the function f to which they are applied. In order to be able to employ different operators simultaneously, however, we must determine relations which exist between the associate functions of the same solution $\psi(\lambda, \theta)$ of (1. 1a), but with respect to different integral operators. In section **5** we determine the associate of P_{21}. This result solves the *initial value problem in the large,* with the data $\psi(\lambda, \theta), \psi_\lambda(\lambda, \theta)$ given on a segment $\lambda = \lambda_0$, $\lambda_0 < 0$, i. e., in the domain of *elliptic* behavior. In section **7** a similar problem is considered with $\psi(0, \theta)$, $\lim_{\lambda \to 0} (-\lambda)^{1/3} \psi_\lambda(\lambda, \theta) = C\psi_{\mathrm{H}}(\mathrm{H}, \theta)_{\mathrm{H}=0}$ (see footnote [22] on p. 468) given, i. e. with data on the transition line, $\lambda = 0$. Our results give information about the behavior of the solution in regions lying in the domain of validity of the corresponding operator. We also clarify the extent to which these results depend upon the coefficient N of equation (1. 1a).

Solutions of linear differential equations with analytic coefficients are analytic functions of two variables. When the ψ's are continued to complex values of the arguments, they represent important subclasses of functions of two complex variables. As indicated in section **8**, the results obtained in sections **5** and **7** can be interpreted from this point of view. They complement in a natural way that portion of the theory of functions of two (and several) complex variables which is devoted to the study of the relations between the coefficients of the series development of a function, its growth and value distribution.

The theorems of sections **5** and **7** illustrate the application of the operator method. The results of the present paper can be employed for the study of various other questions in the theory of differential equations of the type (1.3), in a manner similar to that of the case of regular coefficients.

The author wishes to express his thanks to Dr. Philip Davis for his assistance in the preparation of this paper and various valuable suggestions which have been incorporated.

2. Solution of the initial value problem in the small. In this section we shall, by using a set of particular solutions introduced by the author [7] (cf. [4]), solve the initial value problem *in the small* for equation (1.3) with the data given on the transition line.

The above-mentioned solutions are

$$(2.1)\quad \psi_{n1}(\mathrm{H},\theta)=1!\binom{n}{1}\theta^{n-1}\int_0^{\mathrm{H}}d\mathrm{H} - 3!\binom{n}{3}\theta^{n-3}\int_0^{\mathrm{H}}d\mathrm{H}_3\int_0^{\mathrm{H}}l(\mathrm{H}_2)d\mathrm{H}_2\int_0^{\mathrm{H}_2}d\mathrm{H}+\cdots,$$

$$(2.2)\quad \psi_{n2}(\mathrm{H},\theta)=\theta^n-2!\binom{n}{2}\theta^{n-2}\int_0^{\mathrm{H}}d\mathrm{H}_2\int_0^{\mathrm{H}_2}l(\mathrm{H}_1)d\mathrm{H}+\cdots.$$

In this section, we assume that $l(\mathrm{H})$ is bounded in some interval $-\mathrm{H}'\leq\mathrm{H}\leq\mathrm{H}'$. From this assumption it follows that all the functions ψ_{nk} $(n=1,2,\cdots;k=1,2)$, are twice-differentiable over $[-\mathrm{H}'\leq\mathrm{H}\leq\mathrm{H}';\ -\infty<\theta<\infty]$. These solutions have the further property that

$$(2.3)\qquad \psi_{n1}(0,\theta)=0,\qquad \psi_{n2}(0,\theta)=\theta^n,$$

$$(2.4)\qquad [\partial\psi_{n1}(\mathrm{H},\theta)/\partial\mathrm{H}]_{\mathrm{H}=0}=n\theta^{n-1};\quad [\partial\psi_{n2}(\mathrm{H},\theta)/\partial\mathrm{H}]_{\mathrm{H}=0}=0.$$

THEOREM 2.1. *Let the functions* $\chi_k(\theta)$ $(k=1,2)$, *possess expansions of the form*

$$(2.6)\qquad \chi_k(\theta)=\sum_{\nu=0}^{\infty}a_\nu^{(k)}\theta^\nu,\qquad |\theta|\leq\theta_1.$$

Then

$$(2.7)\qquad \psi(\mathrm{H},\theta)=\sum_{n=1}^{\infty}n^{-1}a_{n-1}^{(2)}\psi_{n1}(\mathrm{H},\theta)+\sum_{n=0}^{\infty}a_n^{(1)}\psi_{n2}(\mathrm{H},\theta)$$

---

[7] These functions were introduced independently in a joint paper [12] (see also [13]), by L. Bers and A. Gelbart, who term them σ-monogenic functions.

converges in a domain

(2. 8) $$D = \mathrm{E}[\,|\,\theta\,| + \sigma\,|\,\mathrm{H}\,| < \theta_1]; \quad \sigma = \sigma(l(\mathrm{H})) > 0.$$

The series (2. 7) *represents a solution of equation* (1. 2) *for which*

(2. 9a) $$\psi(0, \theta) = \chi_1(\theta),$$

(2. 9b) $$[\psi_{\mathrm{H}}(\mathrm{H}, \theta)]_{\mathrm{H}=0} = \chi_2(\theta), \qquad \psi_{\mathrm{H}} = \partial\psi/\partial\mathrm{H},$$

holds.

Proof. From the convergence of the series in (2. 6) it follows that

(2. 10) $$|\,a_n^{(1)}\,| \leq \mu_1/\theta_1^{\,n}; \qquad |\,a_2^{(n)}\,| \leq \mu_2/\theta_1^{\,n}, \qquad (n = 0, 1, 2, \cdots),$$

for some conveniently chosen positive constants μ_1 and μ_2. Inasmuch as $l(\mathrm{H})$ is assumed to be bounded, we have for some positive constant c,

(2. 11) $$|\,l(\mathrm{H})\,| \leq c^2 \qquad \text{for } |\,\mathrm{H}\,| \leq \mathrm{H}'.$$

Now select

(2. 12) $$\sigma > \max\,[\theta_1/\mathrm{H}', c].$$

From (2. 1), (2. 2), and (2. 11), it follows that

(2. 13) $$|\,\psi_{nk}(\mathrm{H}, \theta)\,| \leq (|\,\theta\,| + c\,|\,\mathrm{H}\,|)^n, \qquad \text{for } |\,\mathrm{H}\,| \leq \mathrm{H}'.$$

If $(\mathrm{H}, \theta)\ \varepsilon\ D$, then $|\,\mathrm{H}\,| \leq (\theta_1/\sigma) < \mathrm{H}'$ by (2. 12), so that (2. 13) holds for (H, θ) in this domain *a fortiori*. Thus, for $(\mathrm{H}, \theta)\ \varepsilon\ D$, the series (2. 7) is dominated by

(2. 14) $$(\mu_1 + \mu_2) \sum_{n=0}^{\infty} [(|\,\theta\,| + c\,|\,\mathrm{H}\,|)/\theta_1]^n.$$

But by (2. 12) we have $|\,\theta\,| + c\,|\,\mathrm{H}\,| < \theta_1$ for $(\mathrm{H}, \theta)\ \varepsilon\ D$, and therefore (2. 7) converges absolutely and uniformly in D. By a similar argument, it is clear that the first two derived series also converge absolutely and uniformly in D, and therefore (2. 7) is a solution of (1. 1a). From (2. 3) and (2. 4) it follows that this solution satisfies the initial conditions (2. 9).

Remark. Let A be a class of functions of H and θ which are infinitely differentiable, and which possess the further property that any function of A is uniquely determined by the derivatives $[\partial^{m+n}\psi(\mathrm{H}, \theta)/\partial\mathrm{H}^m\partial\theta^n]_{\mathrm{H}=0,\,\theta=0}$ $(m, n = 0, 1, 2, \cdots)$ (as, for example, the class of functions analytic in a neighborhood of $\mathrm{H} = 0$, $\theta = 0$). If $l(\mathrm{H})$ is infinitely differentiable over $-\mathrm{H}' \leq \mathrm{H} \leq \mathrm{H}'$, and the solution (2. 9) of the partial differential equation

is required to belong to a class A, then it is unique. For since $\psi(0,\theta)=\chi_1(\theta)$, and $\psi_{\mathrm{H}}(0,\theta)=\chi_2(\theta)$, it follows that $(\partial^m\psi(0,\theta)/\partial\theta^m)_{\theta=0}=\chi_1^{(m)}(0)$, and $[\partial^{m+1}\psi(\mathrm{H},\theta)/\partial\mathrm{H}\partial\theta^m]_{\mathrm{H}=0,\theta=0}=\chi_2^{(m)}(0)$ $(m=0,1,2,\cdots)$, while the remaining mixed derivatives may be obtained from $\chi_i(\theta)$ and $l(\mathrm{H})$ by recursion through the repeated differentiation of $\psi_{\mathrm{HH}}+l(\mathrm{H})\psi_{\theta\theta}=0$.

3. Reduction of the equation of mixed type to normal form; the class $\mathfrak{N}$ of differential equations of type (3.1). The function $l(\mathrm{H})$ given by (1.3) will now be assumed to satisfy some additional conditions, the precise formulation of which will be given below. Our aim is to show that certain properties of solutions of (1.1) are independent of the special form of $l(\mathrm{H})$, provided that $l(\mathrm{H})$ satisfies these conditions.

In order to formulate these conditions, it is convenient to reduce (1.1) to the normal form

$$\psi_{\lambda\lambda}+\psi_{\theta\theta}+4N(\lambda)\psi_\lambda=0;\qquad N=(1/8)l^{-3/2}l_{\mathrm{H}},\tag{3.1}$$

by means of the transformation

$$-\lambda(-\mathrm{H})=\int_{t=0}^{-\mathrm{H}}[l(t)]^{-1/2}dt.\tag{3.2}$$

A formal computation shows that $-\lambda(-\mathrm{H})$ has an expansion of the form

$$-\lambda(-\mathrm{H})=(2/3)a_1^{1/2}(-\mathrm{H})^{3/2}+(1/3)a_2a_1^{-1/2}(-\mathrm{H})^{5/2}+\cdots,\tag{3.3}$$

valid in a neighborhood of $\mathrm{H}=0$, while N has an expansion of the form

$$N(\lambda)=(-\lambda)^{-1}[-(1/12)+\sum_{\nu=1}^{\infty}\beta_\nu(-\lambda)^{2\nu/3}],\qquad \beta_1>0,\tag{3.4}$$

valid for $-\lambda_0<\lambda<0$, $\lambda_0>0$. If r_0 is the radius of convergence of the series $\sum_{\nu=1}^{\infty}\beta_\nu x^\nu$, then we have $\lambda_0=r_0^{3/2}$.

We now impose the following requirements[8] on the function $N(\lambda)$. We shall say that a function $N(\lambda)$ of the real variable λ belongs to the class $\mathfrak{N}$, if

(3.5a) $N(\lambda)$ is an analytic function for $-\infty<\lambda<0$, and is real for $\lambda<0$;

[8] These requirements, of course, may be replaced so as to give conditions directly on the function $l(\mathrm{H})$; however, such a formulation would be more involved than that given above.

(3.5b) in the neighborhood of $\lambda = 0$, $N(\lambda)$ possesses an expansion of the form (3.4), valid for $-\lambda_0 < \lambda < 0$; $0 < \lambda_0 \leqq \infty$;

(3.5c) the integral [9] $\exp[-\int_{-\infty}^{\lambda} 2N(t)\,dt] = H(\lambda)$, exists for all $\lambda < 0$.

Under these assumptions, it may be shown that $\lim_{\lambda \to 0^-} (-\lambda)^{1/6} H(\lambda)$ exists.[10] Set S_0 equal to the value of this limit. We shall assume further [11] that

$$(3.5d) \qquad S_0 > 0.$$

Remark. If (3.1) possesses a coefficient $N(\lambda) \varepsilon \mathfrak{N}$, we shall say that the differential equation is of class $\mathfrak{N}$.

Before investigating the general equation (3.1), it is useful to consider the special case

$$(3.6) \qquad l(\mathrm{H}) = -a_1 \mathrm{H}, \qquad a_1 > 0,$$

to which, following [9], p. 862, we shall refer as "the simplified case." In this case $N(\lambda) = (12\lambda)^{-1}$. If the lower limit appearing in the integral in (3.5c) is replaced by any a for which $-\infty < a < 0$, it will be found that conditions (3.5a)-(3.5c) are satisfied. In the simplified case, the generating

---

[9] In compressible fluids this integral equals $l^{-1/4} = (\rho^2(1 - M^2))^{1/4}$.

[10] The existence of S_0 may be established as follows. Select λ_1 such that $-\lambda_0 < \lambda_1 < 0$. Then $H(\lambda) = \exp[\int_{-\infty}^{\lambda_1} (-2N(t))\,dt + \int_{\lambda_1}^{\lambda} (-2N(t))\,dt]$. Introduce

$$T(\lambda) = -(1/6)\log(-\lambda) + \sum_{\nu=1}^{\infty} (6\beta_\nu/(2\nu - 3))(-\lambda)^{2\nu/3},$$

$(-\lambda_0 < \lambda < 0)$; then (independently of λ_1),

$$H(\lambda) = \exp[\int_{-\infty}^{\lambda_1} [-2N(t)]\,dt + T(\lambda) - T(\lambda_1)]$$
$$= \exp[\int_{-\infty}^{\lambda_1} (-2N(t))\,dt - T(\lambda_1)](-\lambda)^{-(1/6)}[1 + \cdots].$$

Thus

$$S_0 = \exp[\int_{-\infty}^{\lambda_1} (-2N(t))\,dt - T(\lambda_1)].$$

It should be pointed out that (4.3) of [9] should be corrected to read as (3.5c) above. Also on p. 860 of [9], line 4 from the bottom, "i" should be omitted. On p. 871, lines 12 and 11, $E_{\bar z}{}^*\dagger$, $E_{\bar z t}{}^*\dagger$, and $E_{\bar z\bar z}{}^*\dagger$ should be replaced by $E_{\bar z}{}^*\dagger$, $E_{\bar z t}{}^*\dagger$, and $E_{\bar z \bar z}{}^*\dagger$. On p. 872, formula (4.9'), the exponent "$-2/3$" should be replaced by "$2/3$."

[11] The value of S_0 in the case of compressible fluids is given in (4.3) of [9]. A detailed derivation of this formula can be found in the appendix p. 898 to the Technical Report No. 10 of the series "Operator Methods in the Theory of Compressible Fluids," Harvard University.

function of the second kind $E\dagger$ of the integral operator is the hypergeometric function of the variable $u = Zt^2/(Z+\bar{Z}) = t^2Z/2\lambda$, $Z = \lambda + i\theta$, $\bar{Z} = \lambda - i\theta$;

$$E_2{}^*\dagger = A_1F(1/6, 5/6, 1/2, u) + B_1u^{1/2}F(2/3, 4/3, 3/2, u), \tag{3.7}$$

and it is possible to represent it in the form

$$E_{21}{}^*\dagger^{(k)} = \sum_{n=0}^{\infty} (t^2Z)^{n-(1/2)+(k/2)} Q\dagger^{(n,k)}(2\lambda) \qquad (k = 1, 2), \tag{3.8}$$

holding for $|Z/2\lambda| < 1$, and

$$E_{22}{}^*\dagger^{(k)} = \sum_{n=0}^{\infty} q\dagger^{(nk)}(2\lambda)/(-t^2Z)^{n-(1/2)+(2k/3)} \qquad (k = 1, 2), \tag{3.9}$$

holding for $|Z/2\lambda| > 1$, where [12]

$$\begin{aligned} &Q\dagger^{(n1)}(2\lambda) \\ &= \frac{(1/6)[(1/6)+1]\cdots[(1/6)+n-1](5/6)[(5/6)+1]\cdots[(5/6)+n-1]}{n!(1/2)[(1/2)+1]\cdots[(1/2)+n-1](2\lambda)^n} \\ &\qquad Q\dagger^{(01)}(2\lambda) = 1, \end{aligned} \tag{3.10}$$

and where [13]

$$\begin{aligned} &q\dagger^{(n1)}(\lambda) \\ &= \frac{(1/6)[(1/6)+1]\cdots[(1/6)+n-1](2/3)[(2/3)+1]\cdots[(2/3)+n-1](-2\lambda}{n!(1/3)[(1/3)+1]\cdots[(1/3)+n-1]} \end{aligned} \tag{3.11}$$

and

$$\begin{aligned} &q\dagger^{(n2)}(\lambda) \\ &= \frac{(5/6)[(5/6)+1]\cdots[(5/6)+n-1](4/3)[(4/3)+1]\cdots[(4/3)+n-1](-2\lambda)^n}{n!(5/3)[(5/3)+1]\cdots[(5/3)+n-1]} \end{aligned} \tag{3.12}$$

4. The integral operator of the second kind for the general equation. In this section we shall show that for the general case there exist two series developments which are analogous to (3.8) and (3.9). They may be written in the form [14]

---

[12] $Q^{(n,3)}(2\lambda)$ has not been listed here inasmuch as the corresponding operator, after integration, yields zero identically.

[13] Formula (5.22) of [9] should be corrected as above. Formula (4.14) of [9] should read "$+ (b_3S_0/(2\Lambda)^{1/6})(2\Lambda/(2\Lambda - t^2(\Lambda+\theta)))^{1/3}F(\cdots$." On p. 880, formula (5.20) should read "for $2|\lambda| < Z$, and $|\lambda|^{2/3} < s_0$."

[14] In addition to developments in powers of u and $(1/u)$, the hypergeometric series has other representations. In section 5 we shall derive representations corresponding to developments in powers of $1-u$ and $1-(1/u)$.

$$E_{21}^{*(k)} \sim \sum_{n=0}^{\infty} (t^2 Z)^{n-(1/2)+k/2} Q^{(nk)}(2\lambda) \qquad (k=1,2),$$
$$(\lambda < 0, |\theta|^2 < 3|\lambda|^2), \tag{4.1}$$

and

$$E_{22}^{*(k)} \sim \sum_{n=0}^{\infty} q^{(nk)}(2\lambda)/(-t^2 Z)^{n-(1/2)+2k/3} \qquad (k=1,2),$$
$$(|\theta|^2 > 3|\lambda|^2, |\lambda| < \lambda_0), \tag{4.2}$$

where [15]

$$(2n+1)Q_\lambda^{(n+1\,1)} + Q_{\lambda\lambda}^{(n1)} + 4FQ^{(n1)} = 0, \qquad Q^{(n+1\,1)}(-\infty) = 0, \tag{4.3}$$
$$Q_\lambda^{(11)} = -4F; \qquad F(\lambda) = -N^2(\lambda) - (1/2)N_\lambda(\lambda)$$
$$Q^{(0k)} = (2\lambda)^{(1/2)-k/2},$$

while the $q^{(nk)}$ are given by the recursion relations

$$q_{\lambda\lambda}^{(0k)} + 4F(\lambda)^{(0k)} = 0 \qquad (k=1,2), \tag{4.4}$$
$$2[n + (2k/3)]q_\lambda^{(nk)} + q_{\lambda\lambda}^{(n+1\,k)} + 4F(\lambda)q^{(n+1\,k)} = 0,$$

and satisfy the initial conditions which are implicit in (4.5).

THEOREM 4.1. *Let $q^{(nk)}(\lambda)$ $(n=0,1,2,\cdots; k=1,2)$, be a set of functions which satisfy the relations* (4.4) *and in the neighborhood of the origin have developments of the form*

$$q^{(nk)}(\lambda) = \sum_{\nu=0}^{\infty} C_\nu^{(nk)}(-\lambda)^{n-(1/2)+(2/3)(k+\nu)}, \tag{4.5}$$

where

$$C_0^{(01)} = 2^{1/6}, \qquad C_0^{(n1)} = q\dagger^{(n1)}(-1), \tag{4.6a}$$

$$C_1^{(n1)} = 0, \tag{4.6b}$$

$$C_0^{(02)} = 2^{5/6}, \qquad C_0^{(n2)} = q\dagger^{(n2)}(-1). \tag{4.6c}$$

Then the series

$$E_{22}^{*(k)} = \sum_{n=0}^{\infty} \frac{q^{(nk)}(\lambda)}{(-t^2 Z)^{n-(1/2)+(2k/3)}} \qquad (k=1,2), \tag{4.7}$$

converges uniformly and absolutely for Z belonging to any closed subdomain of

$$\mathrm{E}[2|\lambda| < |Z|, \lambda \leq 0] \quad \text{and} \quad |t| \geq 1. \tag{4.8}$$

---

[15] The conditions $Q^{(n+1\,1)}(a) = 0$, $a > -\infty$, could also have been imposed.

The series (4.7) *then represents a solution of the partial differential equation*

$$E_{Zt}^{*} - t^2 E_{\bar{Z}t}^{*} + 2Zt\, F E_{Z\bar{Z}}^{*} + 2Zt\, F E^{*} = 0, \tag{4.9}$$

where

$$Z = \lambda + i\theta, \qquad \bar{Z} = \lambda - i\theta, \qquad \lambda = (Z + \bar{Z})/2, \tag{4.10}$$

and is, therefore, when multiplied by $H(\lambda)$, *a generating function of the second kind defined in the domain* (4.8) *for the differential equation* (3.1).

Proof. As has been shown[16] in [9], the series (4.7) converges absolutely and uniformly for $|t| \geq 1$ and for λ, Z belonging to the region $\mathrm{E}[2|\lambda| < |Z|, \lambda < 0, |-\lambda| < \lambda_0]$, and represents a solution of the differential equation (4.9). It remains to show that the convergence holds for $\lambda < -\lambda_0$. We shall first establish several lemmas. It is convenient to introduce the variable $s = (-\lambda)^{2/3}$, noting that $-\lambda$ is positive and choosing the real branch. In this variable, we shall have

$$F(\lambda(s)) = s^{-3} S(s), \tag{4.11}$$

where $S(s)$ is regular for $|s| < |\lambda_0|^{2/3}$. Moreover, in this variable, the relations (4.4) become

$$s^2 q_{ss}^{(0k)} - (1/2) s q_s^{(0k)} + 9S(s) q^{(0k)} = 0, \tag{4.12a}$$

$$\begin{aligned} -3[n + (2k/3)] s^{5/2} q_s^{(nk)} &+ s^2 q_{ss}^{(n+1\,k)} \\ &- (1/2) s q_s^{(n+1\,k)} + 9S(s) q^{(n+1\,k)} = 0. \end{aligned} \tag{4.12b}$$

REMARK. We note that for $q(\lambda)$ and $q(\lambda(s))$ the same symbol q is used.

LEMMA 1. *Let*

$$w^{(k)}(s) = s^{[k-(3/4)]} \sum_{\nu=0}^{\infty} C_\nu^{(0k)} s^\nu \qquad (k = 1, 2). \tag{4.13}$$

Then we have

$$q^{(n+1\,k)}(s) = 3[n + (2k/3)] w^{(k)}(s) \int_0^s [w^{(k)}]^{-2} s_1^{(1/2)} \Big(\int_0^{s_1} w^{(k)} q_{s_2}^{(nk)} ds_2 \Big) ds_1,$$
$$(n = 0, 1, 2, \cdots; k = 1, 2).$$

---

[16] The special case of this theorem where N is that which appears in the theory of compressible fluids, i.e., when it is given by (2.16), has already been stated on p. 878 of [9]. However, the proof given there is only for λ belonging to the domain $\mathrm{E}[2|\lambda| < |Z|, -\lambda_0 < \lambda \leq 0]$. The generalization of this part of the proof to the general case, i.e. $N(\lambda) \in \mathfrak{N}$, is immediate; the main content of the present section is the removal of the restriction $-\lambda_0 < \lambda \leq 0$. For λ_0, see (3.4).

Proof. It has been shown on page 879 of [9] that $w^{(k)}(s) = q^{(0k)}(s)$, so that by (4.12a), $w^{(k)}(s)$ $(k = 1, 2)$ are two independent solutions of the differential equation.

$$s^2 w_{ss} - (1/2) s w_s + 9S(s) w = 0. \tag{4.14}$$

Introduce

$$u^{(n+1\,k)}(s) = 3[n + (2k/3)] \int_0^s [w^{(k)}]^{-2} s_1^{(1/2)} \left(\int_0^{s_1} w^{(k)} q_{s_2}^{(nk)} ds_2 \right) ds_1, \tag{4.15}$$

so that it is necessary to prove that

$$q^{(n+1\,k)}(s) = w^{(k)}(s) u^{(n+1\,k)}(s), \qquad n = -1, 0, 1, 2, \cdots. \tag{4.16}$$

Substituting the right-hand side of (4.16) into (4.12b), we obtain

$$s^2 w^{(k)} u_{ss}^{(n+1\,k)} + u_s^{(n+1\,k)} [2s^2 w_s^{(k)} - (1/2) s w^{(k)}] = 3[n + (2k/3)] s^{5/2} q_s^{(nk)}, \tag{4.17}$$

or

$$w^{(k)} u_{ss}^{(n+1\,k)} + u_s^{(n+1\,k)} [2 w_s^{(k)} - 2^{-1} s^{-1} w^{(k)}] = 3[n + (2k/3)] s^{1/2} q_s^{(nk)}. \tag{4.18}$$

Now it may be easily verified by differentiation that (4.15) is a particular solution of (4.18), and therefore that (4.16) is a solution of (4.12b).

The first terms in the expansion of $w^{(k)}(s) u^{(n+1\,k)}(s)$ are given by

$$\frac{[(3n/2) - (3/4) + k](3n + 2k) C_0^{(nk)}}{[(3n/2) + 2k - (3/2)][(3n/2) + (3/2)]} s^{[(3n/2)+k+(3/4)]}$$
$$+ \Bigg\{ \frac{[(3n/2) + (1/4) + k](3n + 2k) C_1^{(nk)}}{[(3n/2) + 2k - (1/2)][(3n/2) + (5/2)]}$$
$$+ \frac{(3n + 2k) C_1^{(0k)} C_0^{(nk)}}{C_0^{(0k)}} \Bigg[\frac{[(3n/2) - (3/4) + k]}{[(3n/2) - (1/2) + 2k][(3n/2) + (5/2)]}$$
$$- \frac{2[(3n/2) - (3/4) + k]}{[(3n/2) - (3/2) + 2k][(3n/2) + (5/2)]}$$
$$+ \frac{[(3n/2) - (3/4) + k]}{[(3n/2) + (3/2)][(3n/2) + 2k - (3/2)][(3n/2) + (5/2)]} \Bigg] \Bigg\} s^{[(3n/2)+k+(7/4)]} + \cdots.$$

(See (4.13), (4.5) and (4.15)). From the coefficients in this expansion, it may be verified by induction that the conditions (4.6) are satisfied. Thus we have shown that $q^{(n+1\,k)}(s)$ and $w^{(k)}(s) u^{(n+1\,k)}(s)$ are both solutions of (4.12b) and are such that their first two coefficients coincide. Our lemma now follows.

LEMMA 2. *The functions*

$$w_0 = \exp\left[\int_{-\infty}^{\lambda} 2N d\lambda\right] = (1/H), \tag{4.19}$$

(see (3.5c)) *and*

$$w_2 = -2^{11/6}(3S_0H)^{-1}\int_0^{\lambda} H^2 d\lambda, \tag{4.20}$$

$$S_0 = \lim_{\lambda\to 0^-} (-\lambda)^{1/6} \exp\left[-2\int_{-\infty}^{\lambda} N(t)\,dt\right],$$

both satisfy the differential equation (4.14).

Proof. We have

$$\lambda = -s^{3/2}, \qquad d\lambda/ds = -(3/2)s^{1/2}, \qquad d^2\lambda/ds^2 = -(3/4)s^{-1/2}. \tag{4.21}$$

so that

$$dw_0/ds = -3s^{1/2}w_0N, \tag{4.22}$$

and

$$d^2w_0/ds^2 = 9sw_0N^2 + (9/2)sw_0N_\lambda - (3/2)s^{-1/2}w_0N. \tag{4.23}$$

Substituting these values in (4.14) and simplifying them, we obtain

$$2N_\lambda + 4N^2 + 4F = 0, \tag{4.24}$$

which, by condition (4.3), is satisfied identically by N.

To prove that (4.20) is a solution of (4.14), we may proceed as follows. Let

$$w_2 = H^{-1}r. \tag{4.25}$$

Then we obtain the following condition on r in order that (4.14) be satisfied. We have

$$dw_2/ds = (1/H)r_s + r(1/H)_s, \tag{4.26}$$

$$d^2w_2/ds^2 = (1/H)r_{ss} + 2r_s(1/H)_s + r(1/H)_{ss},$$

so that (4.14) becomes

$$s^2[(r_{ss}/H) - 2r_s(H_s/H^2) + r(1/H)_{ss}] \tag{4.27}$$

$$- (s/2)[(r_s/H) + r(1/H)_s] + 9[S(s)](r/H) = 0,$$

or

$$s(r_{ss} - 2r_s(H_s/H)) - (1/2)r_s = 0, \tag{4.28}$$

since $(1/H)$ is a solution of (4.14). The quantity r_s must therefore satisfy

(4.29) $(r_{ss}/r_s) = (2H_s/H) + (1/2s)$, or

$$(d\log(r_s)/ds) = (2d\log H/ds) + 1/2(d\log s/ds),$$

so that $r_s = cH^2s^{1/2}$, and $r = c\int_0^s s^{1/2}H^2ds = -(2/3)c\int_0^\lambda H^2d\lambda$.

LEMMA 3.

$$w^{(2)}(s) = w_2(s), \tag{4.30}$$

$$w^{(1)}(s) = 2^{1/6}S_0w_0(s) + 2^{-2/3}S_1w_2(s), \tag{4.31}$$

where $S_1 = 3\beta_1$, and S_0 has already been given in (3.5c); and hence

$$w^{(1)}(s) = 2^{1/6}S_0H^{-1} + 2^{1/6}S_1S_0^{-1}H^{-1}\int_0^s H^2s^{1/2}ds, \tag{4.32}$$

$$w^{(2)}(s) = 2^{5/6}S_0^{-1}H^{-1}\int_0^s H^2s^{1/2}ds. \tag{4.33}$$

Proof. As we have shown, both sides of (4.30) and (4.31) satisfy the linear differential equation (4.14). Hence, by classical theorems, it suffices to show that the first two coefficients in the expansions coincide. It has been shown in [9], page 870, that in the neighborhood of $\lambda = 0$, $H(\lambda)$ has an expansion of the form

$$H(\lambda) = S_0(-\lambda)^{-1/6} + S_0S_1(-\lambda)^{3/6} + S_0S_2(-\lambda)^{7/6} + \cdots. \tag{4.34}$$

Hence

$$\begin{aligned} &H^2 = S_0^2(-\lambda)^{-1/3} + \cdots, \quad \int_0^\lambda H^2d\lambda = -(3/2)S_0^2(-\lambda)^{2/3} + \cdots, \\ &-2^{11/6}(3S_0H)^{-1}\int_0^\lambda H^2d\lambda = 2^{5/6}(-\lambda)^{5/6} + \cdots. \end{aligned} \tag{4.34a}$$

Therefore, from (4.13), (4.6c) and (4.20), since the series development of $w^{(2)}$ and w_2 begins with $s^{5/4}$, both $w^{(2)}$ and w_2 coincide in the first two coefficients $a_1 = 0$ and $a_2 = 2^{5/6}$ of their series development $a_1(-\lambda)^{1/6} + a_2(-\lambda)^{5/6} + \cdots$. We have, furthermore, from (4.19) and (4.34),

$$S_0w_0 = (-\lambda)^{1/6} - S_1(-\lambda)^{5/6} + \cdots,$$

and from (4.20) and (4.34a) $S_1w_2 = 2^{5/6}S_1(-\lambda)^{5/6} + \cdots$, so that $2^{1/6}S_0w_0 + 2^{-2/3}S_1w_2 = 2^{1/6}(-\lambda)^{1/6} + 0\cdot(-\lambda)^{5/6} + \text{const}(-\lambda)^{9/6} + \cdots$. Therefore, by (4.13), (4.6a) and (4.6b), the first two coefficients of this sum and $w^{(1)}$ coincide.

LEMMA 4. *For an arbitrary $s_2 > 0$ there exist three positive constants $C_1(s_2)$, $C_2(s_2)$, $C_3(s\ )$, such that*

$$(4.35)\quad C_1 s^{k-(3/4)} \leq w^{(k)}(s) \leq C_2 s^{k-(3/4)}, \text{ for } 0 \leq s \leq s_2 < \infty, \quad (k = 1, 2),$$

$$(4.36)\quad | w_s^{(k)} | \leq C_3 s^{k-(7/4)}, \qquad 0 \leq s \leq s_2.$$

Proof. We note that since S_0 and S_1 are both positive, it follows from (4.19), (4.32), and (4.33) that $w^{(k)}(s) \geq 0$, and $w^{(k)}(s) = 0$ only if $s = 0$. By (4.13),

$$(4.37)\quad w^{(k)} s^{-k+(3/4)} = \sum_{\nu=0}^{\infty} C_\nu^{(0k)} s^\nu, \qquad | s | < s_0,$$

so that $w^{(k)} s^{-k+(3/4)}$ is regular in $| s | < s_0$. Since, in addition, the integrals (4.32) and (4.33) are valid for all real $s \geq 0$, the existence of constants for which (4.35) holds, now follows. Again, since

$$(4.38)\quad w_s^{(k)} s^{-k+(7/4)} = \sum_{\nu=0}^{\infty} (\nu + k - (3/4)) C_\nu^{(0k)} s^\nu, \qquad | s | < s_0,$$

$w_s^{(k)} s^{-k+(7/4)}$ is regular in $| s | < s_0$, so that the existence of C_3 with (4.36) follows.

We are now in a position to prove our theorem. Recalling that in (4.15) $u^{(n+1\,k)}(s)$ was defined by

$$(4.15)\quad u^{(n+1\,k)}(s) = 3[n + (2k/3)] \int_0^s [w^{(k)}]^{-2} s_1^{1/2} \left(\int_0^{s_1} w^{(k)} q_{s_2}^{(nk)} ds_2 \right) ds_1,$$

we wish to show that for all s_2, $s_2 > 0$, and for all $\epsilon > 0$, there exists a $C = C(s_2, \epsilon)$ such that

$$(4.39)\quad | u^{(nk)}(s) | \leq C \frac{2^n \Gamma[n + (2k/3)]}{\Gamma(2k/3)\Gamma(n+1)} (1+\epsilon)^n s^{3n/2}, \qquad 0 \leq s \leq s_2.$$

Since, according to page 879 of [9] and (4.13), $w^{(k)}(s) = q^{(0k)}(s)$, we have, by (4.16), $u^{(0k)}(s) = 1$, so that statement (4.39) is obviously true for $n = 0$. From (4.5), (4.13), (4.16), the series development for $u^{(nk)}$ has the form

$$(4.40)\quad u^{(\mu k)}(s) = C_0^{(\mu k)} [C_0^{(0k)}]^{-1} s^{3\mu/2} \left(1 + \sum_{\nu=1}^{\infty} a_\nu^{(\mu k)} s^\nu\right), \qquad 0 \leq s \leq s_0.$$

Let ϵ and s_2 be fixed. In view of (4.40) and the fact that $w^{(k)} \geq 0$ and equals 0 only when $s = 0$, it follows that for every finite n we may determine C so large that the inequality is valid for $\mu = 0, 1, 2, \cdots, n$. For a fixed s_2, choose n so large that

$$(4.41)\quad C_2 C_3 C_1^{-2} [(3n/2) + 2k - (3/2)]^{-1} < \epsilon,$$

where C_1, C_2, C_3 are constants introduced in Lemma 4. Assuming the inequality (4.39) for $\mu \leq n$, we shall show that it holds for $\mu = n+1$. If we integrate the inner integral in (4.15) by parts, and note that $w^{(k)}(0) = 0$, we have, by (4.16),

$$(4.42) \quad u^{(n+1\,k)} = 3[n+(2k/3)] \int_0^s u^{(nk)} s^{(1/2)} ds - 3[n+(2k/3)] \int_0^s (w^k)^{-2} s_1^{1/2} \left(\int_0^{s_1} w^{(k)} w_{s_2}^{(k)} u^{(nk)} ds_2 \right) ds_1.$$

Therefore, using (4.35), (4.36), and (4.39),

$$\begin{aligned} | u^{(n+1\,k)} | &\leq C \frac{3[n+(2k/3)]2^n \Gamma[n+(2k/3)](1+\epsilon)^n}{\Gamma(2k/3)\Gamma(n+1)} \\ &\quad \times \left[\frac{s^{3(n+1)/2}}{(3/2)(n+1)} + \frac{C_2 C_3 s^{3(n+1)/2}}{(3/2)C_1^2[(3n/2)+2k-(3/2)](n+1)} \right] \\ &\leq C \frac{2^{n+1}\Gamma[n+1+(2k/3)](1+\epsilon)^n s^{3(n+1)/2}}{\Gamma(2k/3)\Gamma(n+2)} \\ &\quad \times \left[1 + \frac{C_2 C_3}{C_1^2[(3n/2)+2k-(3/2)]} \right], \end{aligned}$$

and in view of (4.41),

$$| u^{(n+1\,k)} | \leq \frac{2^{n+1}\Gamma[n+1+(2k/3)](1+\epsilon)^{n+1} s^{3(n+1)/2}}{\Gamma(2k/3)\Gamma(n+2)},$$

which is the inequality for $\mu = n+1$.

Finally, we have

$$(4.43) \quad E_{22}^{*(k)} = \sum_{n=0}^{\infty} \frac{w^{(k)} u^{(nk)}}{(-t^2 Z)^{[n-(1/2)+(2k/3)]}},$$

and

$$\left| \sum_{n=N}^{\infty} \frac{w^{(k)} u^{(nk)}}{(-t^2 Z)^{[n-(1/2)+(2k/3)]}} \right| \leq \left| \frac{w^{(k)}}{(-t^2 Z)^{[(2k/3)-(1/2)]}} \right| \sum_{n=N}^{\infty} \left| \frac{u^{(nk)}}{(-t^2 Z)^n} \right|;$$

by (4.39),

$$\sum_{n=N}^{\infty} \left| \frac{u^{(nk)}}{(-t^2 Z)^n} \right| \leq \sum_{n=N}^{\infty} \frac{2^n \Gamma[n+(2k/3)]C}{\Gamma(2k/3)\Gamma(n+1)} \left| \frac{(1+\epsilon)^n s^{(3n/2)}}{(-t^2 Z)^n} \right|,$$

which converges for $| 2\lambda/t^2 Z | < (1+\epsilon)^{-1}$. Since ϵ is arbitrarily small, the theorem is proved.

5. Inversion formulae. In this section we shall determine the inverse of the various integral operators under consideration, i.e., the analytic func-

tion $f(Z)$ which, when substituted in an operator of the form (1.2), yields the given (real) solution $\psi(Z, \bar{Z}) = \boldsymbol{O}(f)$. The determination of the inverse reduces such questions as the determination of the domain of regularity, the determination of the character of the singularities, etc., to the corresponding problems in the theory of analytic functions of one complex variable.

As we stressed in section 1, it is sometimes of interest to consider the integral operator of the first kind with the point of reference *different from the origin.* Such an operator may be written in the form [17]

$$\boldsymbol{p}_1(g, Z_0) = R(Z', \bar{Z}')\psi^*(Z', \bar{Z}'), \tag{5.1}$$

where

$$\psi^*(Z', \bar{Z}') = g(Z') + \sum_{n=1}^{\infty} \frac{(2n)!}{2^{2n} n!} P^{(n)}(Z', \bar{Z}') \int_0^{Z'} \cdots \int_0^{Z'_{n-1}} g(Z'_n) dZ_n \cdots dZ'_1, \tag{5.2}$$

and

$$Z' = Z - Z_0, \qquad \bar{Z}' = \bar{Z} - \bar{Z}_0, \tag{5.3}$$

and the functions $P^{(n)}$ satisfy the recurrence relations,

$$P_{\bar{Z}'}^{(1)} + 2F = 0; \qquad (2n+1)P_{\bar{Z}'}^{(n+1)} + 2P_{Z'\bar{Z}'}^{(n)} + 2FP^{(n)} = 0 \qquad (n = 1, 2, \cdots), \tag{5.4}$$

and are such that

$$P^{(n)}(Z', 0) = 0 \qquad (n = 1, 2, \cdots). \tag{5.5}$$

Here $F = F[(1/2)(Z' + \bar{Z}') + (1/2)(Z_0 + \bar{Z}_0)]$ is the function introduced in (4.3), where $\lambda = (1/2)(Z + \bar{Z})$, see (1.1a).

THEOREM 5.1. *Let $\psi(Z, \bar{Z})$ be a real solution of* (1.1a) *which is regular in the neighborhood of $Z = Z_0$, $\bar{Z} = \bar{Z}_0$, and is such that*

$$2\psi(Z, \bar{Z}) = \boldsymbol{p}_1(g; Z_0) + \overline{\boldsymbol{p}_1(g; Z_0)}. \tag{5.6}$$

Then

$$2\psi(Z, \bar{Z}_0) = g(Z') + \bar{R}(0, Z')\bar{g}(0), \quad 2\psi(Z_0, \bar{Z}) = R(0, \bar{Z}')g(0) + \bar{g}(\bar{Z}'). \tag{5.7}$$

Proof. This follows immediately from (5.1), (5.2) and (5.5), by the substitution of $\bar{Z}' = 0$ and $Z' = 0$, respectively.

---

[17] (5.1) is somewhat altered form of P(f), see (1.2). In (5.13) the relation between the corresponding f and g is given.

Similarly, we now consider the integral operator of the second kind with an arbitrary reference point Z_0. Such an operator is defined by [18]

$$(5.8) \qquad \mathrm{P}_{21}(f;Z_0) = \int_{-1}^{+1} E_{21}(Z',\bar{Z}',t) f[(Z'/2)(1-t^2)]dt/(1-t^2)^{1/2}$$

(see (1.2)), where

$$(5.9) \qquad E_{21}(Z',\bar{Z}';t) = H(2\lambda')[1+\sum_{n=1}^{\infty}(t^2Z')^n Q^{(n)}(2\lambda')],$$

$$2\lambda = Z'+\bar{Z}'; \quad 2\lambda_0 = Z_0+\bar{Z}_0; \quad H(2\lambda') = \exp[-\int_{-\infty}^{2\lambda'} N(\lambda+\lambda_0)d\lambda],$$

and where the functions $Q^{(n)}(2\lambda')$ are given by the recursion formula

$$(5.10) \qquad (2n+1)[dQ^{(n+1)}/d\lambda'] + [d^2Q^{(n)}/d\lambda'^2] + 4F(\lambda'+\lambda_0)Q^{(n)} = 0,$$
$$[dQ^{(1)}/d\lambda'] + 4F(\lambda'+\lambda_0') = 0,$$

and are such that

$$(5.11) \qquad Q^{(n)}(\lambda') = 0, \text{ for } \lambda' = 0, \qquad (n = 1, 2, \cdots).$$

Remark. The domain of regularity R of this operator is given by $(\lambda-\lambda_0)^2+\theta^2 < 4\lambda^2$. This may be established by the method of [7], pp. 36-39, replacing λ by $\lambda' = \lambda - \lambda_0$ throughout. The operator (5.8) may be written in the alternate form

$$(5.12) \qquad \boldsymbol{p}_{21}(g;Z_0) = H(2\lambda')\psi^*(Z',\bar{Z}'),$$

where $\psi^*(Z'\bar{Z}')$ has the same form as that given in (5.2), except that $P^{(n)}(Z',\bar{Z}')$ are to be replaced by $Q^{(n)}(2\lambda')$. Furthermore, f and g are related by

$$(5.13) \qquad g(Z') = \int_{-1}^{+1} f[(Z'/2)(1-t^2)]dt/(1-t^2)^{(1/2)},$$

THEOREM 5.2. *Let there be given a (real) solution*

$$\psi(Z',\bar{Z}') = \Psi(\lambda',\theta') = H(2\lambda')\psi^*(Z',\bar{Z}') = H(2\lambda')\Psi^*(\lambda',\theta')$$
$$= (1/2)(\boldsymbol{p}_{21}(g;Z_0) + \overline{\boldsymbol{p}_{21}(g;Z_0)}),$$

where

$$g(Z) = \sum_{n=0}^{\infty} \mu_n Z^n, \qquad \mu_n = \mu_{n1} + i\mu_{n2}.$$

[18] The operator P_{21} with the reference point $Z_0 = 0$ has been considered in section 4.

Furthermore, let

$$(5.14)\qquad \Psi^*(0,\theta') = \sum_{n=0}^{\infty} \beta_n \theta'^n,$$

$$\Psi^*_{\lambda'}(0,\theta') = [H(0)]^{-1}\Psi_{\lambda'}(0,\theta') - H_{\lambda'}(0)[H(0)]^{-2}\Psi(0,\theta') = \sum_{n=0}^{\infty} \gamma_n \theta'^n.$$

Then μ_n *are related to* β_n *and* γ_n *as follows:*

$$(5.15)\qquad \mu_{2n1} = (-1)^n\beta_{2n}, \quad \mu_{2n+1\,2} = (-1)^{n+1}\beta_{2n+1} \qquad (n = 0, 1, 2, \cdots),$$

$$(5.16)\qquad \mu_{2n+2-k\,k} = \begin{vmatrix} D_0^{(k)} & D_1^{(k)} & \cdots & D_{n-1}^{(k)} & D_n^{(k)} \\ a_{00}^{(k)} & a_{10}^{(k)} & \cdots & a_{n-1\,0}^{(k)} & a_{n0}^{(k)} \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & \cdots & a_{0\,n-1}^{(k)} & a_{n\,n-1}^{(k)} \end{vmatrix} \qquad \begin{matrix}(n = 0, 1, 2, \cdots; \\ k = 1, 2),\end{matrix}$$

where $a_{0\nu} = -1$, $\nu = 0, \cdots, n-1$, *and where*

$$(5.17\text{a})\qquad D_n^{(1)} = (-1)^n(2n+1)^{-1}\gamma_{2n} - [(2n+1)!]^{-1}\sum_{j=1}^{n}(-1)^{n-j}C_{2j}A_{2j}[(2n-2j)!]\beta_{2n-2j},$$

$$(5.17\text{b})\qquad a_{nm}^{(1)} = -[(2n+1)!]^{-1}C_{2n-2m-1}A_{2n-2m-1}(2m+1)!,$$

$$(5.17\text{c})\qquad C_n = (2n)!\,2^{-2n}(n!)^{-1}, \qquad A_n = Q^{(n)}(\lambda')\Big|_{\lambda=0},$$

and

$$(5.18\text{a})\qquad D_n^{(2)} = (-1)^n 2^{-1}n^{-1}\gamma_{2n-1} - [(2n)!]^{-1}\sum_{j=1}^{n-1}C_{2j}A_{2j}[(2n-2j-1)!](-1)^{n-j}\beta_{2n-2j-1},$$

$$(5.18\text{b})\qquad a_{nm}^{(2)} = -[(2n)!]^{-1}C_{2n-m-1}A_{2n-m-1}(2m)!.$$

Proof.

$$(5.19)\qquad 2\Psi^*(\lambda',\theta') = \sum_{n=0}^{\infty}\mu_n(\lambda'+i\theta')^n + C_1Q^{(1)}(\lambda')\sum_{n=0}^{\infty}(n+1)^{-1}\mu_n(\lambda'+i\theta')^{n+1} + \cdots + \sum_{n=0}^{\infty}\bar{\mu}_n(\lambda'-i\theta')^n + C_1Q^{(1)}(\lambda')\sum_{n=0}^{\infty}(n+1)^{-1}\bar{\mu}_n(\lambda'-i\theta')^{n+1} + \cdots,$$

and therefore

$$(5.20)\qquad 2\Psi^*(0,\theta') = \sum_{n=0}^{\infty}(\mu_{n1}[(i)^n + (-i)^n] + \mu_{n2}[(i)^{n+1} + (-i)^{n+1}])\theta'^n = 2\sum_{n=0}^{\infty}\beta_n\theta'^n.$$

Equations (5.15) follow immediately from this. Furthermore,

$$(5.21)\quad 2[d\Psi^*(\lambda',\theta')/d\lambda'] = \sum_{n=0}^{\infty} n\mu_n(\lambda'+i\theta')^{n-1} + C_1Q^{(1)}(\lambda')\sum_{n=0}^{\infty}\mu_n(\lambda'+i\theta')^n$$
$$+ C_1[dQ^{(1)}(\lambda')/d\lambda']\sum_{n=0}^{\infty}(n+1)^{-1}\mu_n(\lambda'+i\theta')^{n+1}+\cdots$$
$$+\sum_{n=0}^{\infty} n\bar{\mu}_n(\lambda'-i\theta')^{n-1} + C_1Q^{(1)}(\lambda')\sum_{n=0}^{\infty}\bar{\mu}_n(\lambda'-i\theta')^n$$
$$+ C_1[dQ^{(1)}(\lambda')/d\lambda']\sum_{n=0}^{\infty}(n+1)^{-1}\bar{\mu}_n(\lambda'-i\theta')^{n+1}+\cdots$$

Whence, by (5.11) and (5.17d),

$$(5.22)\quad 2[d\Psi^*(\lambda',\theta')/d\lambda']\Big|_{\lambda'=0} = \sum_{n=0}^{\infty} n\mu_n(i)^{n-1}\theta'^{n-1}$$
$$+ C_1A_1\sum_{n=0}^{\infty}(n+1)^{-1}(i)^{n+1}\mu_n\theta'^{n+1}$$
$$+ C_2A_2\sum_{n=0}^{\infty}(n+1)^{-1}(n+2)^{-1}(i)^{n+2}\mu_n\theta'^{n+2}+\cdots$$
$$+\sum_{n=0}^{\infty} n\bar{\mu}_n(-i)^{n-1}\theta'^{n-1} + C_1A_1\sum_{n=0}^{\infty}(n+1)^{-1}(-i)^{n+1}\bar{\mu}_n\theta'^{n+1}$$
$$+ C_2A_2\sum_{n=0}^{\infty}(n+1)^{-1}(n+2)^{-1}(-i)^{n+2}\bar{\mu}_n\theta'^{n+2}+\cdots$$
$$=\sum_{n=0}^{\infty}\gamma_n\theta'^n.$$

From this it follows that

$$(5.23)\quad \mathrm{Re}\{i^n[(n+1)\mu_{n+1} + C_1A_1n^{-1}\mu_{n-1} + C_2A_2[n(n-1)]^{-1}\mu_{n-2} + \cdots + C_nA_n\mu_0(n!)^{-1}]\} = \gamma_n \qquad (n=0,1,\cdots).$$

It follows from this that

$$(5.24)\quad \gamma_{2n}(-1)^n = (2n+1)\mu_{2n+1\,1} + [(2n)!]^{-1}\sum_{j=1}^{2n}(2n-j)!C_jA_j\mu_{2n-j\,1}$$

$$(5.25)\quad \gamma_{2n+1}(-1)^{n+1} \qquad (n=0,1,\cdots)$$
$$= (2n+2)\mu_{2n+2\,2} + [(2n+1)!]^{-1}\sum_{j=1}^{2n+1}(2n+1-j)!C_jA_j\mu_{2n+1-j\,2} \qquad (n=0,1,\cdots).$$

From the first of these relations and (5.15), we obtain

$$\mu_{2n+1\,1} = \sum_{j=0}^{n-1} a_{nj}^{(1)} \mu_{2j+1\,1} + D_n^{(1)} \qquad (n = 0, 1, \cdots), \tag{5.26}$$

while from the second, we obtain

$$\mu_{2n2} = \sum_{j=0}^{n-1} a_{nj}^{(2)} \mu_{2j} + D_n^{(2)} \qquad (n = 0, 1, \cdots). \tag{5.27}$$

The expressions (5.16) now follow.

Theorem 5.2 gives us the coefficients of the associate of the second kind in terms of the initial values χ_1, χ_2. The connection between the coefficients β_n, γ_n on one side, and μ_n on the other, can be interpreted as a projection in the coefficient space. While the associate of the integral operator of the first kind of $\psi = \Sigma A_{mn} z^m \bar{z}^n$, $A_{nm} = \bar{A}_{mn}$, can be determined in terms of the A_{n0}'s *independently* of the equation, the above-mentioned projection depends upon the equation, since the elements of the determinant (5.16) depend on the $Q^{(n)}$'s. (See (5.17b), (5.17c)).

Our result represents much more than the solution of the initial value problem. All properties of the associate g (a function of a complex variable), or f (see (5.13)), can be "translated" into theorems on the behavior of ψ in the domain R (cf. the remark on p. 461) of the regularity of the operator. For details, see [6], p. 299. As an example, it is of interest to formulate the results which follow from the Hadamard theory of coefficients of entire or meromorphic functions. If the coefficients of f satisfy the well-known Hadamard conditions, so that f has in a domain $B \subset R$, poles as its only singularities, then ψ will, at the corresponding points, have certain branch points of order $n + (1/2)$, n integer. If $f = f_1/f_2$, where f_1 and f_2 are functions which are regular in B, the representation

$$\psi(Z', \bar{Z}') = \int_{t=-1}^{1} E_{21}(Z', \bar{Z}', t) f(Z'(1-t^2)/2) dt/(1-t^2)^{1/2}, \tag{5.28}$$

holds in B.

Remark. ψ is an infinitely many-valued function possessing branch point at the point p_ν where f_2 has zeros, and we have to specify what branch of it is represented by (5.28).

If on the other hand g satisfies the Hadamard conditions, so that we can write $g = (g_1/g_2)$, where g_1 and g_2 are regular in B, then we can write

$$(5.29)\quad \psi(Z',\bar{Z}') = -(2\pi)^{-1}\int_{t=-1}^{1} E_{21}(Z',\bar{Z}',t)$$
$$\times\int_{\tau=-1}^{1} g_1[Z'(1-\tau^2)(1-t^2)/2]/g_2[Z'(1-\tau^2)(1-t^2)/2](1-t^2)^{-1/2}\tau^{-2}d\tau dt$$

or [19]

$$\psi(Z',\bar{Z}') = H(\lambda')[(g_1/g_2)$$
$$+\sum_{n=1}^{\infty}[2^n(n!)^2]^{-1}(2n)!Q^{(n)}(2\lambda')\int_0^{Z'}(Z'-\zeta)^{n-1}(g_1(\zeta)/g_2(\zeta))d\zeta].$$

The above formulas can be used for the study of the growth and value distribution in the neighborhood of a point of accumulation of singularities of ψ. Thus, the theory of integral operators gives us insight into the structure of the singularities of a solution of a differential equation.

In this connection it will be of interest to clarify the possibility of the representation of *single-valued* solutions of (1.1) in the neighborhood of a point of accumulation of single-valued singularities.

It is well known that the only singularities of linear differential equations of elliptic type with analytic coefficients which are single-valued in the real plane are fundamental solutions, say $\psi^{(L1)}(Z,Z_0) = A(\log\zeta + \log\bar{\zeta}) + B$, $A \equiv A(Z,\bar{Z};Z_0,\bar{Z}_0)$, $B \equiv B(Z,\bar{Z};Z_0,\bar{Z}_0)$, $\zeta = Z - Z_0$, $\bar{\zeta} = \bar{Z} - \bar{Z}_0$, and their derivatives with respect to the parameters Z_0, $\bar{Z}_0$. (See [8], p. 473). Let Z_ν be an infinite set of points in a domain B. The question arises whether there exists an infinite sum

$$\phi = \sum_{\nu=1}^{\infty}[\psi^{(L1)}(Z,Z_\nu) + H_\nu(Z)],$$

of such singularities with conveniently chosen solutions H_ν's, which are regular in B and possess the property that $D(\phi) = \int\int_B \phi^2 w\, dx\, dy < \infty$. Here w is some (positive) weight function. In order to answer this question, we can repeat the considerations carried out in [2], p. 143-147, where we introduced fundamental solutions which are normalized with respect to the domain B. (They possess the property that they are orthogonal to every regular solution ϕ with finite $D(\phi)$ of the differential equation). The considerations of [2] can be repeated in this case, and we get necessary and sufficient conditions for the existence of functions ϕ, $D(\phi) < \infty$, possessing the above singularities at the prescribed points Z_ν. If sufficient conditions are satisfied, the above considerations will also give us a representation of the function ϕ in terms

[19] If g_2 is a polynomial, the above integral can be evaluated in terms of elementary (i.e., algebraic and logarithmic) functions.

of the kernel function of a complete orthonormal system with a weight function w.

6. Other representations of the integral operator of the second kind. In the simplified case, in addition to the representation (3.8) for P_{21} and (3.9) for P_{22}, we can obtain two further representations of (3.7) in powers of $v = 1 - u$ and $(u-1)/u$. We shall show that similar developments hold in the general case.

The function [20] $E^* = E/H$, considered as a function of

$$(6.1) \qquad u = t^2 Z/2\lambda, \quad \text{and} \quad s = (-\lambda)^{2/3},$$

satisfies the differential equation

$$(6.2) \quad u(1-u)E^*_{uu} + (1/2 - 2u)E^*_u - (5/36)E^* + (2s/3)(2u-1)E^*_{us} + 9^{-1}s(2+3u^{-1})E^*_s - (4/9)s^2 E^*_{ss} - 4(S(s) - (5/144))E^* = 0.$$

If we now introduce

$$(6.3) \qquad \tilde{E}^* = E^* u^{-1/2},$$

we find that $\tilde{E}^*$ satisfies the differential equation

$$(6.4) \quad u(1-u)\tilde{E}^*_{uu} + ((3/2) - 3u)\tilde{E}^*_u - [4S(s) + (8/9)]\tilde{E}^* + (8/9)s\tilde{E}^*_s + (2s/3)(2u-1)\tilde{E}^*_{us} - (4s^2/9)\tilde{E}^*_{ss} = 0.$$

and upon introducing the variable

$$(6.5) \qquad v = 1 - u = 1 - t^2 Z(2\lambda)^{-1},$$

we have

$$(6.6) \quad v(1-v)\tilde{E}^*_{vv} + [(3/2) - 3v]E^*_v - [4S(s) + (8/9)]\tilde{E}^*_s + (8s/9)\tilde{E}^*_s + (2s/3)(2v-1)\tilde{E}^*_{vs} - (4s^2/9)\tilde{E}^*_{ss} = 0.$$

It will be observed that equations (6.4) and (6.6) are identical, except that the role of u and v is interchanged. It therefore follows that (6.6) has solutions of the form

$$(6.7) \qquad \tilde{E}^* = v^{-1/2}[(-\lambda)^{(k/2)-(1/2)} + \sum_{n=1}^{\infty} (-2\lambda v)^n Q^{(nk)}(2\lambda) \qquad (k = 1, 2),$$

and

$$(6.8) \qquad \tilde{E}^* = v^{-1/2} \sum_{n=0}^{\infty} [[q^{(nk)}(\lambda)](-2\lambda v)^{-[n-(1/2)+(2k/3)]}] \qquad (k = 1, 2),$$

whence, in the original variables,

$$(6.9) \quad E^* = tZ^{1/2}(2\lambda - t^2 Z)^{-1/2}[(-\lambda)^{(k/2)-(1/2)} + \sum_{n=1}^{\infty} (t^2 Z - 2\lambda)^n Q^{(n)}(2\lambda)] \qquad (k = 1, 2),$$

[20] In this section we consider both E_{I} and E_{II}, but the subscripts have been omitted.

valid for $|(t^2Z - 2\lambda)/2\lambda| < 1$, and

$$(6.10) \qquad E^* = tZ^{1/2}[(2\lambda - t^2Z)^{-1/2} \sum_{n=0}^{\infty} [q^{(nk)}(\lambda)](t^2Z - 2\lambda)^{-[n-(1/2)+(2k/3)]}] \qquad (k = 1, 2),$$

valid for $|(t^2Z - 2\lambda)/2\lambda| > 1$.

7. On solutions of the initial value problem in the large. One of our aims is to consider the problem of the solution of the initial value problem in the large. More precisely: if on a segment of the transition line, the value of a solution and of its derivative in a direction normal to the line are given, how far can the solution be continued into the elliptic region, what are the locations and the character of its singularities, etc.? In the present section, we shall obtain a partial answer to these questions. The properties thus obtained are, to a certain extent, independent of the function $F(\lambda)$ (or N). We shall now obtain a sufficient condition which will guarantee the existence of a solution $\psi(\lambda, \theta)$ of equation (3.1) which is regular in a subregion B of the elliptic region, and which satisfies given initial data on the transition line $\lambda = 0$. The problem is reduced to the determination of the domain of regularity of an analytic function of a complex variable in terms of the coefficients of its Taylor development. This question can be answered by a number of results in the theory of analytic functions, and we shall not enter into it here.

Let $\mathfrak{l}$ be a simple curve which connects the points -1 and $+1$ of the *complex* t-plane and lies completely in $1 \leq |t| \leq A < \infty$, A sufficiently large. For each point ζ, we denote by $c(\zeta; \mathfrak{l})$ the curve traced out by the values $(1/2)\zeta(1 - t^2)$ as t traverses $\mathfrak{l}$ from -1 to $+1$. Furthermore, we denote by $T_{\mathfrak{l}}(B)$ the domain

$$(7.1) \qquad T_{\mathfrak{l}}(B) = \sum_{\zeta \varepsilon B} c(\zeta; \mathfrak{l}).$$

If, for instance, $\mathfrak{l}$ is the upper half of the unit circle, then $c(\zeta; \mathfrak{l})$ is the circle $|z - \zeta/2| = |\zeta/2|$, so that $T_{\mathfrak{l}}(B)$ consists of the union of all circles passing through the origin and points $\mathfrak{P} \varepsilon B$, and having the segment $\widehat{0\mathfrak{P}}$ as diameter. More generally, as t traverses an arbitrary curve from -1 to $+1$, the expression $(\zeta/2)(1 - t^2)$ varies on a closed curve c which passes through the origin. Thus $T_{\mathfrak{l}}(B)$ consists of the union of closed curves $c(\zeta, \mathfrak{l})$, each passing through the origin, and each obtained from the fixed curve c by a dilation of amount $|\zeta|$ and a rotation of amount $\arg \zeta$. It is clear that any closed curve passing through the origin may be selected as the fixed curve c, and that the origin will be a boundary point for all domains $T_{\mathfrak{l}}(B)$.

THEOREM 7.1. *Let $\chi_1(\theta)$ and $\chi_2(\theta)$ be two functions of θ which possess representations of the form*

$$(7.2)\qquad \chi_j(\theta) = \sum_{\nu=0}^{\infty} a_\nu^{(j)}\theta^\nu, \qquad 0 \leq \theta \leq \theta_1, \theta_1 > 0, (j = 1, 2).$$

Let a function $f(\zeta)$ *be defined by*

$$(7.3)\qquad f(\zeta) = \zeta^{1/6} \sum_{\nu=0}^{\infty} c_\nu \zeta^\nu,$$

where

$$(7.4)\qquad c_\nu = -(-2i)^{1/6}[3^{1/2}\pi S_0{}^2 \operatorname{Im}(A_2\bar{A}_1)]^{-1} \times [-a_\nu^{(1)}\bar{d}_0 J_\nu^{(1)} + \sum_{k=1}^{2}(-1)^k a_\nu^{(k)}\bar{d}_k J_\nu^{(2)}],$$

with $\operatorname{Im}[A_2\bar{A}_1] \neq 0$, *and* [21]

$$(7.5)\qquad d_0 = -(2/3)i^{3/2}S_0A_2, \quad d_1 = -(2^{5/3}/3)i^{1/6}S_0S_1A_1 \quad d_2 = -i^{1/6}S_0A_1.$$

$$J_\nu^{(1)} = (1/2)e^{-(2/3)n\pi i}(e^{-(2/3)\pi i} - 1)[\Gamma(-1/3)\Gamma(\nu+1)/\Gamma(\nu+2/3)] \qquad (n = 0, \pm 1),$$

$$(7.6)\qquad J_\nu^{(2)} = (1/2)e^{-(4/3)n\pi i}(e^{-(4/3)\pi i} - 1)[\Gamma(1/3)\Gamma(\nu+1)/\Gamma(\nu+4/3)] \qquad (n = 0, \pm 1).$$

If a domain B *situated in* $\mathrm{E}[3^{1/2} \mid \lambda \mid < \theta, \theta > 0, \lambda \leq 0]$, *contains the segment* $0 \leq \theta \leq \theta_1$, *and if* $f(\zeta)$ *is regular in* $T_{\mathrm{I}}(B)$, *then*

$$(7.7)\qquad \psi(\lambda, \theta) = \operatorname{Im}[\int_{\mathrm{I}} E(Z, \bar{Z}, t) f[Z(1-t^2)/2](1-t^2))^{-1/2}\,dt],$$

where $E = A_1E^{(1)} + [Z(1-t^2)/2]^{2/3}A_2E^{(2)}$, *is a solution of the differential equation* (2.3), *which is regular in* B *and is such that* [22]

$$(7.8a)\qquad \lim_{\lambda\to 0^-} \psi(\lambda, \theta) = \chi_1(\theta),$$

$$(7.8b)\qquad \lim_{\lambda\to 0^-} (-\lambda)^{1/3}[\partial\psi(\lambda, \theta)/\partial\lambda] = \chi_2(\theta).$$

Proof. With t on I and $Z \varepsilon B$, by Theorem 4.1, $E(Z, \bar{Z}, t)$ is regular. By assumption, $f[Z(1-t^2)/2]$ is also regular there. Hence, (7.7) is a solution of the differential equation (3.1). The boundary relations (7.8a) and (7.98) follow [22] from Theorem 6.1 of [9].

---

[21] The three independent possibilities in (7.6), are covered by $n = 0$, $n = 1$, $n = -1$, respectively.

[22] It follows from (3.3) that $\lim_{\lambda\to 0^-} (-\lambda)^{-1/3}d\lambda/d\mathrm{H} = (3a_1/2)^{1/3}$, so that the limit in (7.8b) essentially coincides with $\lim_{\mathrm{H}\to 0} \psi(\mathrm{H}, \theta)$.

It should be further noted that the initial value problem for the equation $\psi_{\mathrm{HH}} + (-\mathrm{H})^s\psi_{\theta\theta} = 0$, $s > -1$, has been studied in the papers [10], [15], [18], and formulas for the representation of the solution have been obtained for these particular cases.

THEOREM 7.2. *Let $\chi_1(\theta)$ and $\chi_2(\theta)$ be two real functions of θ which are regular in $0 \leq \theta \leq \theta_1$, and which possess expansions of the form* (7.2). *If the function $g(\zeta) = \zeta^{-1/6} f(\zeta) = \sum_{\nu=0}^{\infty} c_\nu \zeta^\nu$, where c_ν are the linear combinations of the $a_\nu^{(j)}$ given by* (7.4), *possesses at the point $\zeta = \alpha$, $(4/3)\pi < \arg \alpha < 2\pi$, an expansion of the form*

$$g(\zeta) = \zeta + \sum_{\nu=0}^{\infty} a_\nu (\zeta - \alpha)^{-\nu}, \tag{7.9}$$

valid for $0 < |\zeta - \alpha| < \infty$, then there exists a solution $\psi(\lambda, \theta)$ of the differential equation (3.1) *which is regular in $\mathrm{E}[3^{1/2} |\lambda| < |\theta|, \theta > 0, |\lambda| \leq 0]$, and such that*

$$\lim_{\lambda \to 0^-} \psi(\lambda, \theta) = \chi_1(\theta), \tag{7.10a}$$

$$\lim_{\lambda \to 0^-} (-\lambda)^{1/3} \psi_\lambda(\lambda, \theta) = \chi_2(\theta), \tag{7.10b}$$

and which possesses a singularity at the point of infinity.

Proof. We consider the function

$$\psi(\lambda, \theta) = \mathrm{Im}[\int_{\mathrm{I}} E(Z, \bar{Z}, t) f[Z(1-t^2)/2] dt/(1-t^2)^{1/2}] \\ = \mathrm{Im}[\mathrm{P}_{22}(f)], \tag{7.11}$$

where I is the upper half of the unit circle, and

$$E(Z, \bar{Z}, t) \\ = \{\exp[-\int_{-\infty}^{2\lambda} 2N(t) dt]\} [A_1 E^{*(1)} + [Z(1-t^2)/2]^{2/3} A_2 E^{*(2)}]. \tag{7.12}$$

For $Z \varepsilon R \equiv \mathrm{E}[3^{1/2} |\lambda| < |\theta|, \theta > 0, \lambda \leq 0]$ and $t \varepsilon \mathrm{I}$, it is clear that the points $\zeta = Z(1-t^2)/2$ will be confined to the sector $0 \leq \arg \zeta \leq (4/3)\pi$. Hence if α is confined to the sector $(4/3)\pi \leq \arg \zeta < 2\pi$, and $g(\zeta)$ possess the expansion (7.9), then $f[Z(1-t^2)/2]$ will be regular for $Z \varepsilon R$, $t \varepsilon \mathrm{I} - 0$ (the endpoints, $t = \pm 1$ give no trouble in the integration). Thus, by the preceding theorem, $\psi(\lambda, \theta)$ is a solution of (3.1) which is regular in $\mathrm{E}[3^{1/2} |\lambda| < \theta, 0 < \theta < \infty, \lambda \leq 0]$, and is such that the initial value relations (4.10) are valid.

Now $f(\zeta) = \zeta^{7/6} + \zeta^{1/6} \sum_{\nu=0}^{\infty} a_\nu (\zeta - \alpha)^{-\nu} = \zeta^{7/6} + \zeta^{1/6} h(\zeta)$, where $h(\zeta) = O(1)$ for $\zeta \to \infty$.

Also $\psi(\lambda, \theta) = \mathrm{Im}[\mathrm{P}_{22}(f)] = \mathrm{Im}[\mathrm{P}_{22}(\zeta^{7/6})] + \mathrm{Im}[\mathrm{P}_{22}(\zeta^{1/6} h(\zeta))]$. We shall show that on the line $\lambda = 0$ the first integral becomes infinite, while the second remains bounded. From (7.5), (7.6) and (7.7), it is clear that on $\lambda = 0, E(Z, \bar{Z}, t)$ becomes $E(i\theta, -i\theta, t) = \mathrm{const}(-it^2\theta)^{-1/6}$. Thus

$P_{22}(\zeta^{7/6})\Big|_{\lambda=0} = \text{const}\,(i\theta/2)\int_{I}[(1-t^2)^{1/2}(-t)^{-1/3}]dt$. Depending upon the determination of the root, we have [23]

$$\int_{I}[(1-t^2)^{1/2}(-t)^{-1/3}]dt = (-1)^{2/3}e^{-(4/3)n\pi i}(e^{-(4/3)\pi i}-1)$$
$$[\Gamma(1/3)\Gamma(5/6+4/6)][\Gamma(1+5/6)]^{-1} \quad (n=0,\pm 1)$$

(see [9], p. 889). Thus $\operatorname{Im}[P_{22}(\zeta^{7/6})]\Big|_{\lambda=0} = e_1\theta,\ e_1 \neq 0$, so that

$$\lim_{\theta\to\infty}[\operatorname{Im} P_{22}(\zeta^{7/6})]\Big|_{\lambda=0} = \infty.$$

Furthermore,

$$\operatorname{Im}[P_{22}(\zeta^{1/6}h(\zeta))]\Big|_{\lambda=0} = \operatorname{Im}\int_{I}(-t)^{-1/3}(1-t^2)^{-1/2}h(i\theta(1-t^2)/2)\,dt.$$

Thus

$$|\operatorname{Im}[P_{22}(\zeta^{1/6}h(\zeta))]\Big|_{\lambda=0} \leq \int_{I}|(-t)^{-1/3}(1-t^2)^{-1/2}h(i\theta(1-t^2)/2)|\,|dt| = O(1),$$
$$\text{for } \theta\to\infty.$$

In (2.1) and (2.2) of section 2, we have considered a set of particular solutions ψ_{nk} (introduced in [4]). In [7], the operator R_2 was introduced [24] transforming the analytic function $\sum_{n=0}^{\infty}(\alpha_n+i\beta_n)\zeta^n$, $\zeta=\theta+iH$ as follows:

(7.13) $\quad R_2[\Sigma(\alpha_n+i\beta_n)\zeta^n] = \Sigma[\alpha_n(\theta+iH)^{[n]} + i\odot\beta_n(\theta+iH)^{\{n\}}]$.

Here $\operatorname{Im}(\theta+iH)^{[n]} = \psi_{n1}$, $\operatorname{Im}(\theta+iH)^{\{n\}} = \psi_{n2}$. For the symbol $\odot$ see [8], p. 455. It is of interest that the operator $\operatorname{Im} R_2$ coincides in certain cases with the integral operator $\operatorname{Re} P_{22}$ considered in section 4.

THEOREM 7.3. *Let $\chi_1(\zeta)$ and $\chi_2(\zeta)$ be regular for $|\zeta|<\zeta_0$, and let $l(H)$ be regular at $H=0$. Then*

(7.14) $\quad \operatorname{Im} R_2[\int_0^{\zeta}\chi_2(t)dt + i\chi_1(\zeta)]$

$$= \operatorname{Re}\int_{t=-1}^{+1}E_{22}(\lambda,\theta,t)f(Z(1-t^2)/2))dt/(1-t^2)^{1/2},$$

where $f(\zeta)$ is given by

(7.15) $\quad f(\zeta) = -\{(-2i\zeta)^{1/6}/3^{1/2}\pi S_0{}^2\operatorname{Im}[A_2\bar{A}_1]\}$

$$\times[-\bar{d}_0\int_{C_2}t^{-1/3}\chi_1(\sigma)dt - \bar{d}_1\int_{C_2}t^{-5/3}\chi_1(\sigma)dt$$
$$+(2/3a_1)^{1/3}\bar{d}_2\int_{C_2}t^{-5/3}\chi_2(\sigma)dt], \qquad \sigma = -2i\zeta(1-t^2),$$

and where the constants appearing here have been defined in (7.5).

[23] See footnote 21.

[24] In [7] the operator R_2 was designated by P_2.

Proof: It is easily shown that

$$(7.16)\quad \operatorname{Im} R_2\Big[\int_0^{\zeta} \chi_2(t)\,dt + i\chi_1(\zeta)\Big] = \sum_{n=1}^{\infty} (1/n)\, a_{n-1}^{(2)}\psi_{n1}(H,\theta) + \sum_{n=0}^{\infty} a_n^{(1)}\psi_{n2}(H,\theta) = \psi(H,\theta),$$

where $\chi_k(\theta) = \sum_{\nu=0}^{\infty} a^{(k)}\theta^{\nu}$ $(k = 1, 2)$.

If we assume that $l(H)$ is an analytic function of H, it follows from Theorem 7.1 that $\psi(H,\theta)$ is a solution of the differential equation (1.3) which is analytic in H and θ in a sufficiently small neighborhood of, say, $N = [-\theta_1 \leq \theta \leq \theta_1, H_0 \leq H \leq 0]$, $\theta_1 > 0$, and is such that $\psi(0,\theta) = \chi_1(\theta)$, $\partial\psi(H,\theta)/\partial H\big|_{H=0} = \chi_2(\theta)$. By Theorem 6.1 of [9], the expression $\operatorname{Re}\int_{-1}^{1} E_{22}(\lambda,\theta;t) f(Z(1-t^2)/2)(1-t^2)^{-1/2}dt$ is also a solution, possessing the initial values $\psi(0,\theta) = \chi_1(\theta)$ and $\partial\psi(H,\theta)/\partial H)\big|_{H=0} = \chi_2(\theta)$ on the line $H = 0$. We shall now show that this solution must also be regular in the neighborhood N. From this it will follow by the uniqueness theorem (cf. the remark on p. 449) that the two solutions considered must be identical.

It is clear, first of all, that the function

$$\int_{-1}^{+1} E_{22}(\lambda,\theta;t) f(Z(1-t^2)/2)(1-t^2)^{-1/2}dt$$

must be regular in $s^{1/2}$, where $s = (-\lambda)^{2/3}$ (see (5.7) of [9]). Since by (3.3), H is a regular function of s in the vicinity of $s = 0$, it suffices to show that this integral must be regular in s.

If we replace in the differential equation (1.1a) the variable λ by s, we obtain

$$(7.17)\quad 4s^{-1}(\partial^2\psi/\partial s^2) + 9(\partial^2\psi/\partial\theta^2) = 24s^{-1}[\beta_1 + \beta_2 s + \cdots](\partial\psi/\partial s) = 0.$$

Since the above solution is known to be regular in $s^{1/2}$, we have

$$(7.18)\qquad \psi = \sum_{n=0}^{\infty} a_n s^n + \gamma_p s^{(p+1)/2} + \gamma_{p+1} s^{(p+3)/2} + \cdots,\ p \text{ even}, \geq 2.$$

In view of the fact that the fractional powers do not combine with the integer powers, we need merely substitute the second portion of (7.16) into (7.17). When this is done, and when we compare the coefficient of $s^{(p-5)/2}$, we obtain $4((p+1)/2)((p-1)/2)\gamma_p = 0$. Hence $\gamma_p = 0$, and this shows (by induction) that fractional powers cannot appear in a solution of the form (7.18).

8. Connection between the results obtained by the operator method and the general theory of functions of two complex variables. It is of considerable interest that the results obtained by the operator method can be considered as complementing an important portion of the general theory of analytic functions of several complex variables. While in the case of functions of one complex variable, the methods of Nevanlinna, Ahlfors, and others establish relations between the growth of the function and value distribution in the whole plane, generalizations of these methods (by exploiting the idea of the distinguished boundary surface and that of the extended class) lead us to relations between the growth and value distribution in certain (non-analytic) *three*-dimensional manifolds of the (four diemnsional) space of two complex variables. These investigations in two variables can be complemented by results establishing relations between coefficients of the series development, say of an entire function, and its growth, so that from properties of the coefficients we can draw conclusions about the growth and the value distribution in certain three-dimensional manifolds of the z_1, z_2-space.

As it has been mentioned in [2], one of the aims of the theory of functions of two complex variables is to investigate the relations between the properties of coefficients, growth and value distribution in certain non-analytic *two*-dimensional manifolds, e. g., in the real plane $y_1 = 0$, $y_2 = 0$.

The fact that a function $f = \sum_{m,n} a_{mn} z^m \bar{z}^n$ of two complex [25] variables $z, \bar{z}$ satisfies a linear differential equation with *entire* coefficients,

$$A = \sum_{\nu,\mu} A_{\nu\mu} z^\nu \bar{z}^\mu, \qquad B = \sum_{\nu,\mu} B_{\nu\mu} z^\nu \bar{z}^\mu, \qquad F = \sum_{\nu,\mu} C_{\nu\mu} z^\nu \bar{z}^\mu,$$

can be expressed by requiring that the a_{mn} satisfy infinitely many relations

$$\begin{aligned}(8.1)\quad (m+n)(n+1)a_{m+1\, n+1} &+ \sum_{\nu=1}^{m} \sum_{\mu=0}^{n} A_{\nu\mu}(m-\nu+1)a_{m-\nu+1\, n-\mu} \\ &+ \sum_{\nu=0}^{m} \sum_{\mu=0}^{n} (n-\mu-1) B_{\nu\mu} a_{m-\nu,\, n-\mu+1} + \sum_{\nu=0}^{m} \sum_{\mu=0}^{n} C_{\nu\mu} A_{m-\nu\, n-\mu} = 0 \\ &(m = 0, 1, 2, \cdots; n = 0, 1, 2, \cdots),\end{aligned}$$

where $A_{\nu\mu}$, $B_{\nu\mu}$, $C_{\nu\mu}$ go to zero in a certain way as $\nu \to \infty$, $\mu \to \infty$.

On the other hand, the class of functions which we obtain in this manner has a number of interesting properties which sharply distinguish it from other analytic functions of two complex variables. The method of integral operators indicates that this class of functions has very many properties in common with

[25] We note that when $f(z, \bar{z})$ is continued to complex values of x and y, z and $\bar{z}$ are two *independent* variables.

the class of functions $\psi(z, \bar{z}) = f_1(z) + f_2(\bar{z})$, i. e., sums of two functions of *one* variable. For instance, these functions can have singularities only on the surfaces $z = \text{const.}$, $\bar{z} = \text{const.}$ If, on the other hand, our function satisfies a differential equation with *singular* coefficients, say (3. 1), with N satisfying the conditions indicated in section 3, then the character of these solutions, when continued to complex values of the arguments, changes considerably.[26] Thus these functions of two complex variables, in addition to the surfaces $z = \text{const.}$, $\bar{z} = \text{const.}$, can have singularities on the singularity surface of the coefficients, in our case on $z + \bar{z} = 0$.

These results can be generalized to the study of *pairs* of functions. If we consider a pair of solutions ψ_1, ψ_2 of the differential equation (1. 1) (or (3. 1)) in the real space, it is natural to pair the two real solutions whose series developments have the form

$$\psi_1 = \sum_{n=0}^{\infty} \rho^n (a_n \cos n\phi - b_n \sin n\phi) + \sum_{\substack{m=0 \\ m \neq n}}^{\infty} \sum_{n=0}^{\infty} \rho^m (c_{mn}^{(1)} \cos n\phi + d_{mn}^{(1)} \sin n\phi),$$

$$\psi_2 = \sum_{n=0}^{\infty} \rho^n (b_n \cos n\phi + a_n \sin n\phi) + \sum_{\substack{m=0 \\ m \neq n}} \sum_{n=0} \rho^m (c_{mn}^{(2)} \cos n\phi + d_{mn}^{(2)} \sin n\phi),$$

$$z = \rho e^{i\phi}, \qquad \bar{z} = \rho e^{-i\phi}.$$

(We note that the coefficients a_n and b_n can be given arbitrarily, while the coefficients $c_{mn}^{(k)}$, $d_{mn}^{(k)}$ $(k = 1, 2)$, are uniquely determined by the requirements that ψ_k satisfies th equation (1. 1) or the relation (8. 1).

The theory of integral operators has been applied in some cases ([6], pp. 321-330, and [14]) to obtain theorems on value distribution in the real plane $y_1 = 0$, $y_2 = 0$, and leads to similar results in the case of other non-analytic surfaces. The theory of integral operators indicates that the above theorems represent only special cases of relations which exist for functions whose coefficients satisfy relation of the type indicated in (8. 1). In particular, from the integral representation of functions satisfying a differential equation of order $2n$, $n > 1$, follow similar theorems for solutions of linear differential equations of higher order. The author hopes in the future to return to the investigations of these questions in greater detail.

DIVISION OF ENGINEERING SCIENCES,
HARVARD UNIVERSITY.

---

[26] These considerations may also be regarded as a generalization of investigations of solutions of ordinary differential equations with singular coefficients (Fuchs theory) to the case of two variables, in which case a partial differential equation arises instead of an ordinary one.

BIBLIOGRAPHY

[1] Ludwig Baumgarten, "Beiträge zur Theorie der ganzen Funktionen von zwei komplexen Veränderlichen," *Monatshefte für Mathematik und Physik*, vol. 25 1914), pp. 3-70.

[2] Stefan Bergman, "Sur les fonctions entières et meromorphes de deux variables complexes," *Compositio Mathematica*, vol. 3 (1936), pp. 136-173.

[3] ———, "Zur Theorie der Funktionen, die eine lineare partielle Differentialgleichung befriedigen, I," *Recueil Mathematique*, new series, vol. 2 (1937), pp. 1169-1198.

[4] ———, "A formula for the stream function of certain flows," *Proceedings of the National Academy of Sciences*, vol. 29 (1943), pp. 276-281.

[5] ———, "Linear operators in the theory of partial differential equations," *Transactions of the American Mathematical Society*, vol. 53 (1943), pp. 130-155.

[6] ———, "Certain classes of analytic functions of two real variables and their properties," *Transactions of the American Mathematical Society*, vol. 57 (1945), pp. 299-331.

[7] ———, "On two-dimensional flows of compressible fluids," *N. A. C. A. Technical Note 972* (1945).

[8] ———, "Two-dimensional subsonic flows of a compressible fluid and their singularities," *Transactions of the American Mathematical Society*, vol. 62 (1947), pp. 452-498.

[9] ———, "Two-dimensional transonic flow patterns," *American Journal of Mathematics*, vol. 70 (1948), pp. 856-891.

[10] ———, "An initial value problem for equations of mixed type," *Bulletin of the American Mathematical Society*, vol. 56 (1949), pp. 165-174.

[11] ———, "On tables for the determination of transonic flow patterns," *Hans Reissner Anniversary Volume*, Brooklyn Polytechnic Institute (1949), pp. 13-36.

[12] L. Bers and A. Gelbart, "On a class of differential equations in mechanics of continua," *Quarterly of Applied Mathematics*, vol. 1 (1943), pp. 168-188.

[1]3 ———, "On a class of functions defined by partial differential equations," *Transactions of the American Mathematical Society*, vol. 56 (1944), pp. 67-93.

[14] Herman Chernoff, "Complex solutions of partial differential equations," *American Journal of Mathematics*, vol. 68 (1946), pp. 455-478.

[15] E. Holmgren, "Sur un probleme aux limites pour l'equation $y^m \partial^2 z/\partial x^2 + \partial^2 z/\partial y^2 = 0$," *Arkiv fur Mathematik, Astronomi och Fysik*, vol. 19, B (1927), pp. 1-3.

[16] Jules Sire, "Sur les fonctions entières de deux variables d'ordre apparent total fini," *Rendiconti del Circolo Matematico di Palermo*, vol. 31 (1910), pp. 1-91.

[17] ———, "Sur les fonctions entières de deux variables d'ordre apparent total fini et a croissance regulière par rapport a l'une des variables," *Journal de Mathématiques pures et appliqués*, vol. 9 (1913), pp. 1-37.

[18] F. Tricomi, "Sulle equazioni lineari alle derivato parziali di 2° ordina di tipo misto," *Memorie della Reale Accademia Nazionale dei Lincei*, vol. 14 (1923), pp. 133-247.

ON THE APPLICABILITY OF LINEAR DIFFERENTIAL OPERATORS OF INFINITE ORDER TO FUNCTIONS OF CLASS $L^2(B)$.*

By Philip Davis.

1. Introduction. In the plane of the complex variable $z = x + iy$, we shall designate by $L^2(B)$ the class of functions $f(z)$ which are regular and single valued in a fixed domain B and such that

$$\int\int_B |f(z)|^2 dxdy < \infty. \tag{1}$$

By introducing the integral

$$(f, g) = \int\int_B f\bar{g}\, dxdy \tag{2}$$

as an inner product, and $\| f \|^2 = (f, f)$ as a norm, the class $L^2(B)$ becomes a Hilbert space. This space is intimately related to problems in conformal mapping, and has been studied extensively. Reference is made to Bergman [2] where a systematic theory will be found.

In the present paper we study the class of bounded linear functionals defined on $L^2(B)$ considered as differential operators of infinite order. We show that such an operator is generated by an entire function of exponential type, and is a member of a class of functions, designated by $E^2(B)$ which are transforms of the functions of $L^2(B)$ arising from a Laplace kernel. The properties of $E^2(B)$ are investigated, and there is obtained a generalization of a theorem due to Muggli [7] on the applicability of linear differential operators where the case of functions analytic in a circle has been considered. In later sections, the inversion of this transform is discussed, and these considerations are applied to the question of the completeness of sets $\{e^{\lambda_n z}\}$ for $L^2(B)$, and to the coefficient problem for functions of exponential type with a preassigned indicator. This extends a familiar theorem of Lindelöf.

We shall find it convenient to recall a few basic facts about the space $L^2(B)$. To every B, there corresponds a complete orthonormal system of functions $\{\phi_n(z)\}$, and a "strong" Riesz-Fischer Theorem holds. Moreover, there is a function of two complex variables $K_B(z, \bar{t})$, known as the Bergman

* Received January 22, 1951; revised July 12, 1951.

kernel of the domain, which has proved to be of prime importance in the theory. The kernel belongs to $L^2(B)$ in both z and $\bar{t}$, and may be defined by

$$(3) \qquad K_B(z, \bar{t}) = \sum_{n=0}^{\infty} \phi_n(z)\,(\phi_n(t))^-$$

but is independent of the particular choice of complete orthonormal system used in (3). Here $(\phi_n(t))^-$ designates the complex conjugate of $\phi_n(t)$.

In the present paper we shall restrict our domains B as follows. It will be required that B be schlicht, bounded, and simply connected. Moreover, if the boundary of B is b, it will be required that all the points of the plane which do not lie in $B + b$ form a single region whose boundary is precisely b. Thus a slit circle would be ruled out. It has been shown by Farrell [5] that for such domains, the set of powers is complete. If, therefore, B is such a domain, there exists a unique complete set of orthonormal polynomials

$$(4) \qquad p_n(z) = \sum_{k=0}^{n} a_{nk} z^k; \qquad \operatorname{Re}(a_{nn}) \geqq 0,\; a_{nn} \neq 0,$$

where the coefficients a_{nk} have been determined by the familiar Gram-Schmidt orthonormalization process. Although a number of our theorems go through for more general domains, we shall find it convenient to restrict B as above so that a complete orthonormal set of polynomials is available.

In addition to the class $L^2(B)$, we shall be dealing with linear differential operators of infinite order

$$(5) \qquad L(d) = \sum_{n=0}^{\infty} \beta_n d^n,$$

where the symbolic d^n is thought of as having the force of an n-th derivative evaluated at the origin: $L(d)f = \sum_{n=0}^{\infty} \beta_n f^{(n)}(0)$. It is convenient to replace the d in (5) by the complex variable z and to think of the operator $L(d)$ as being generated by the formal power series [1] $L(z) = \sum_{n=0}^{\infty} \beta_n z^n$. Our initial aim is to distinguish a class of operators (5) which will be applicable to functions of $L^2(B)$, and to characterize their generating power series.

2. On the applicability of linear differential operators. At the outset, we make the following definition. Suppose that $L(d)$ is given by (5), and that $f(z) \in L^2(B)$ has the Fourier expansion

[1] On the utility of this device for the discussion of certain problems arising in interpolatory function theory, see, e. g., Boas [3], Buck [4].

$$(6) \qquad f(z) = \sum_{n=0}^{\infty} \alpha_n p_n(z), \qquad \alpha_n = (f, p_n).$$

We shall say that $L(d)$ is *applicable* (B) to f if and only if the series

$$(7) \qquad \sum_{n=0}^{\infty} \alpha_n \sum_{k=0}^{n} \beta_k p_n^{(k)}(0) = \sum_{n=0}^{\infty} \alpha_n \sum_{k=0}^{n} a_{nk}\beta_k k!$$

converges. If the sum in (7) is r, then we shall write $L(d)_B f = r$.

Let us observe that if $p(z)$ is a polynomial of degree N, then $p(z) = \sum_{n=0}^{N} \alpha_n p_n(z)$, so that

$$(8) \qquad L(d)_B p = \sum_{n=0}^{N} \alpha_n \sum_{k=0}^{n} a_{nk}\beta_k k! = \sum_{k=0}^{N} \beta_k \sum_{k=n}^{N} \alpha_n a_{nk} k!$$
$$= \sum_{k=0}^{\infty} \beta_k p^{(k)}(0) = L(d)p.$$

Moreover, if $L(z)$ has a positive radius of convergence, and if $p(z) = \sum_{n=0}^{N} c_n z^n$, then we shall also have

$$(9) \qquad L(d)_B p = \sum_{k=0}^{\infty} \beta_k p^{(k)}(0) = \sum_{k=0}^{\infty} \beta_k c_k k! = \sum_{k=0}^{N} c_k L^{(k)}(0) = p(d)L(z).$$

By virtue of this observation, we may say that $L(d)$ is applicable (B) to $f(z)$ if and only if $\sum_{n=0}^{\infty} \alpha_n L(d) p_n(z)$ converges.

THEOREM 1. *$L(d)$ is applicable (B) to all functions of $L^2(B)$ if and only if*

$$(10) \qquad \sum_{n=0}^{\infty} \Big| \sum_{k=0}^{n} a_{nk}\beta_k k! \Big|^2 < \infty.$$

In this case, the operator is bounded.

Proof. By our definition, the series $\sum_{n=0}^{\infty} \alpha_n \sum_{k=0}^{n} a_{nk}\beta_k k!$ is required to converge for all sequences $\{\alpha_n\}$ with $\sum_{n=0}^{\infty} |\alpha_n|^2 < \infty$. A familiar theorem of Landau [6] tells us that (10) is a necessary and sufficient condition for this. If we set

$$(11) \qquad K^2 = \sum_{n=0}^{\infty} \Big| \sum_{k=0}^{n} a_{nk}\beta_k k! \Big|^2$$

then, by the Schwarz inequality, $|L(d)_B f|^2 \leqq K^2 \sum_{n=0}^{\infty} |\alpha_n|^2 = K^2 \| f \|^2$, establishing the boundedness of the operator.

For the circle $C_r: |z| < r$, the orthonormal polynomials are given explicitly by

$$p_n(z) = \pi^{-\frac{1}{2}}(n+1)^{\frac{1}{2}} z^n / r^{n+1} \qquad (n = 0, 1, \cdots), \tag{12}$$

and as a result of their simple structure, we see from (7) that if $L(d)$ is applicable to $f(z) \,\varepsilon\, L^2(C_r)$, then $L(d)_{C_r} f = L(d) f$. Thus for the case of the circle, applicability as defined by (7) reduces to ordinary applicability (cf. also (49)). For the circle, the criterion (10) becomes

$$\sum_{n=0}^{\infty} |\beta_n|^2 n!(n+1)!r^{-2n} < \infty. \tag{13}$$

In his connection, cf. Muggli [7].

For a fixed domain B, we shall designate the class of power series $L(z)$ whose coefficients $\{\beta_n\}$ satisfy (10) by $E^2(B)$. It will appear shortly that the power series in this class have more than formal meaning; indeed, they are entire functions of exponential type. The exact scope of the linear differential operators of class $E^2(B)$ is indicated by the following converse of Theorem 1.

THEOREM 2. *Any bounded linear functional defined on $L^2(B)$ may be represented as a linear differential operator of infinite order and of class $E^2(B)$.*

Proof. From the general theory (see, e. g., Stone [9], p. 62) it is known that if L is such a functional, there will exist an $h(z) \,\varepsilon\, L^2(B)$ such that

$$Lf = (f, h) \qquad \text{for all } f \,\varepsilon\, L^2(B). \tag{14}$$

Suppose that $h(z)$ has the Fourier expansion $h(z) = \sum_{n=0}^{\infty} \alpha_n p_n(z)$. Set up the infinite system of linear equations in the unknowns β_k $(k = 0, 1, \cdots)$,

$$\sum_{k=0}^{n} a_{nk}\beta_n k! = \alpha_n \qquad (n = 0, 1, \cdots). \tag{15}$$

Inasmuch as the system (15) is triangular, and $a_{nn} \neq 0$, it will be possible to arrive at a solution in a step-by-step fashion. Since $\sum_{n=0}^{\infty} |\alpha_n|^2 < \infty$, the power series $L(z) = \sum_{k=0}^{\infty} \beta_k z^k$ has the property (10). Thus, $L(z) \,\varepsilon\, E^2(B)$, and $L(d)$ is applicable (B) to all $f(z) \,\varepsilon\, L^2(B)$. If now $f(z) = \sum_{n=0}^{\infty} \gamma_n p_n(z)$, then $L(d)_B f(z) = \sum_{n=0}^{\infty} \gamma_n \sum_{k=0}^{n} a_{nk}\beta_k k!$. But

$$(f, h) = \iint_B \sum_{n=0}^{\infty} \gamma_n p_n(z) \left(\sum_{m=0}^{\infty} \alpha_m p_m(z) \right)^- dxdy$$

$$= \sum_{n=0}^{\infty} \gamma_n \alpha_n = \sum_{n=0}^{\infty} \gamma_n \sum_{k=0}^{n} a_{nk} \beta_k k! = L(d)_B f.$$

We next give an explicit construction for the function $h(z)$ associated with the functional $L(z)$ through the relation (14). It is obtained by operating on one variable of the kernel $K_B(z, \bar{t})$ by $L(d)$. In this connection, see also Aronszajn [1]. We shall merely state the result.

THEOREM 3. *Let $L(z) \varepsilon E^2(B)$. Then the function*

$$h(t) = \sum_{n=0}^{\infty} (L(d) p_n(z))^- p_n(t) = (L(d_z)_B K_B(z, \bar{t}))^- \tag{16}$$

belongs to $L^2(B)$, and is such that

$$L(d)_B f = (f, h) \tag{17}$$

for all $f \varepsilon L^2(B)$. The subscript in $L(d_z)$ indicates that the operation is to be performed on the variable z.

The quantity $(h, h)^{\frac{1}{2}}$ is the norm of the operator, and we have explicitly

$$\| L(d) \|^2 = \sum_{n=0}^{\infty} | L(d) p_n(z) |^2 = \sum_{n=0}^{\infty} \left| \sum_{k=0}^{n} a_{nk} \beta_k k! \right|^2. \tag{18}$$

The norm of $L(d)$ depends, of course, upon the domain B, and when we wish to emphasize the role of B we shall write $\| L(d) \|_B$.

3. The class $E^2(B)$. In this section we characterize more closely the class of power series $E^2(B)$ whose Maclaurin coefficients $\{\beta_n\}$ are subject to (10). First we shall find a representation for functions of this class, showing that it is identical with the class of transforms of $f \varepsilon L^2(B)$ arising from a Laplace kernel.

THEOREM 4. *$g(z) \varepsilon E^2(B)$ if and only if*

$$g(z) = \iint_B e^{zw} (f(w))^- \, dudv; \qquad w = u + iv, \tag{19}$$

with $f(z) \varepsilon L^2(B)$.

Proof. Let $f(z) \varepsilon L^2(B)$ and have Fourier coefficients $\alpha_n = (f, p_n)$.

Then we have

$$g(z) = \int\int_B e^{zw}(f(w))^- dudv = \int\int_B \sum_{n=0}^{\infty} z^n w^n (f(w))^-/n!\, dudv$$
$$= \sum_{n=0}^{\infty} \beta_n z^n,$$

where $\beta_n = (n!)^{-1}(w^n, f(w))$. Therefore,

$$\sum_{k=0}^{n} a_{nk}\beta_k k! = \sum_{k=0}^{i} a_{nk}(w^n, f(w)) = (p_n, f) = \bar{\alpha_n},$$

so that condition (10) holds. Conversely, given a $g(z) = \sum_{n=0}^{\infty} \beta_n z^n$ with (10) holding, there exists by the Riesz-Fischer Theorem an $f(z) \in L^2(B)$ such that the quantities $(\sum_{k=0}^{n} a_{nk}\beta_k k!)^-$ are its Fourier coefficients:

$$\sum_{k=0}^{n} a_{nk}\beta_k k! = (p_n, f) \qquad (n = 0, 1, \cdots). \tag{20}$$

We now claim that $g(z) = \sum_{k=0}^{\infty} \beta_k z^k = \int\int_B e^{zw}(f(w))^- dudv$. To show this, it suffices, as above, to prove that

$$\beta_k k! = (w^k, f(w)) \qquad (k = 0, 1, \cdots). \tag{21}$$

From (20), we have

$$\sum_{k=0}^{n} a_{nk}\beta_k k! = (\sum_{k=0}^{n} a_{nk} w^k, f(w)) = \sum_{k=0}^{n} a_{nk}(w^k, f(w)), \tag{22}$$

$(n = 0, 1, \cdots)$. This triangular system may now be solved step-by-step, and yields (21) as its solution.

It will be convenient to abbreviate the transform (19) as $g = \mathfrak{L}_B(f)$, and when this notation is employed, it will be understood that $f \in L^2(B)$. We shall now characterize the growth of the transforms $\mathfrak{L}_B(f)$, but before doing so, it will be useful to recall a number of definitions and theorems which are of importance in the theory of entire functions of exponential type. For this material, reference is made to Pólya [8]. By the Phragmén-Lindelöf indicator of an entire function of exponential type $f(z)$ in the direction θ is meant the quantity

$$h(\theta) = \limsup_{r\to\infty} r^{-1} \log |f(re^{i\theta})| \tag{23}$$

If we have $h(\theta) \leqq \sigma$ for $0 \leqq \theta \leqq 2\pi$, we shall speak of an entire function

of exponential type at most σ, a class designated by $(1,\sigma)$. If $f(z)=\sum_{n=0}^{\infty}a_nz^n$ is an entire function of exponential type, then the associated series

$$\mathfrak{B}(f)=\sum_{n=0}^{\infty}a_n n!/z^{n+1}, \tag{24}$$

is known as the Borel transform of f. An entire function $f(z)$ is of class $(1,\sigma)$ if and only if the radius of convergence of $\mathfrak{B}(f)$ is at most σ. The closed convex hull of the set of singularities of $\mathfrak{B}(f)$, known as the conjugate indicator diagram of f, will be designated by $D(f)$. If $f\,\varepsilon\,(1,\sigma)$, then it possesses a representation of the form

$$f(z)=(2\pi i)^{-1}\int_C e^{zw}\mathfrak{B}f(w)\,dw, \tag{25}$$

where C is any simple closed curve containing $D(f)$ in its interior. If B is a bounded convex domain, its supporting function $k(\theta)$ is defined by

$$k(\theta)=\max_{(x,y)\,\varepsilon\,\bar{B}}(x\cos\theta+y\sin\theta). \tag{26}$$

The indicator of an entire function of exponential type and the supporting function of its conjugate indicator diagram are related intimately by the formula

$$h(\theta)=k(-\theta) \qquad 0\leqq\theta\leqq 2\pi. \tag{27}$$

THEOREM 5. *Let $\hat{B}$ designate the closed convex hull of a given domain B. Every function $g(z)$ of the class $E^2(B)$ is an entire function of exponential type. If $h(\theta)$ is its indicator, and if $k(\theta)$ is the supporting function of $\hat{B}$, then*

$$h(\theta)\leqq k(-\theta), \qquad 0\leqq\theta\leqq 2\pi. \tag{28}$$

Furthermore, we have $D(g)\subseteq B$.

Proof. With $g(z)\,\varepsilon\,E^2(B)$, we have, by Theorem 4, $g=\mathfrak{L}_B(f)$ for some $f\,\varepsilon\,L^2(B)$. Applying the Schwarz inequality, we obtain

$$\begin{aligned} |g(re^{i\theta})|^2 &\leqq \|f\|^2\int\int_B|\exp(re^{i\theta}w)|^2\,du\,dv \\ &= \|f\|^2\int\int_B \exp\{2r(u\cos\theta-v\sin\theta)\}\,du\,dv. \end{aligned} \tag{29}$$

Now for $(u,v)\,\varepsilon\,\hat{B}$, and hence for $(u,v)\,\varepsilon\,B$, we have by (26)

$$u\cos\theta-v\sin\theta\leqq k(-\theta), \qquad 0\leqq\theta\leqq 2\pi,$$

so that (29) becomes $|g(re^{i\theta})|^2\leqq\|f\|^2\exp\{2rk(-\theta)\}$ area (B). Therefore,

$h(\theta) = \limsup_{r\to\infty} r^{-1} \log |g(re^{i\theta})| \leqq k(-\theta)$. Finally, by (27), $h(-\theta)$ is the supporting function of $D(g)$. By virtue of the last inequality, we have $h(-\theta) \leq k(\theta)$, and this implies that $D(g) \subseteq B$.

While $D(g) \subseteq B$, the former need not be interior to the latter. We shall give an example which shows that in certain cases the boundary of $\hat{B}$ may contain a boundary point of $D(g)$. Take B as the unit circle. For this domain $\bar{B} = \hat{B}$, and $(z^m, z^n) = \pi\delta_{mn}/(n+1)$ $(m, n = 0, 1, \cdots)$. Consider the function $\phi(z) = \pi^{-1} \sum_{n=0}^{\infty} z^n/(n+1)^{\frac{1}{2}}$. ϕ is regular in the unit circle, and $\| \phi \|^2 = \sum (n+1)^{-2} < \infty$ (cf. (12)), so that $\phi \in L^2(B)$. Now $\psi(z) = \mathfrak{L}_B(\phi)$ $= \sum_{n=0}^{\infty} z^n/n!(n+1)^{3/2}$, so that $\mathfrak{B}(\psi) = \sum_{n=0}^{\infty} (n+1)^{-3/2} z^{-n-1}$. This last series converges at points exterior to B, and $\mathfrak{B}(\psi)$ possesses a singularity on the circumference of B.

A basis for the space $E^2(B)$ is provided by the functions

$$(30) \qquad q_n(z) = \iint_B e^{zw} (p_n(w))^- \, dudv \qquad (n = 0, 1, \cdots).$$

By Theorem 5, the q's are entire of exponential type. Since also

$$e^{zw} = \left(\sum_{k=0}^{n-1} + \sum_{k=n}^{\infty}\right) z^k w^k / k!,$$

and each $p_n(z)$ is orthogonal to all polynomials of lower degree, we have

$$(31) \qquad q_n(z) = b_{nn} z^n/n! + b_{n+1n} z^{n+1}/(n+1)! + \cdots \qquad (n = 0, 1, \cdots).$$

It is easily shown that all functions of the class $E^2(B)$ may be expanded in a series of q's, convergent uniformly in every bounded domain.

4. Inversion formulas for the $\mathfrak{L}_B$ transform. In the present section, we derive a number of formulas for inverting the $\mathfrak{L}_B$ transform (19).

THEOREM 6. *Let* $g = \mathfrak{L}_B(f)$. *Then*

$$(32) \qquad f(z) = \sum_{n=0}^{\infty} (g(d) p_n(z))^- p_n(z),$$

$$(33) \qquad f(z) = \sum_{n=0}^{\infty} (g(d_w)_B K_B(w, \bar{z}))^-,$$

$$(34) \qquad f(z) = \sum_{n=0}^{\infty} p_n(z) \{ (2\pi i)^{-1} \int_C g(z) \mathfrak{B}(p_n(z)) \, dz \}^-,$$

where C is any simple closed curve containing the origin in its interior. If it is assumed, in addition, that $D(g)$ lies in the interior of B, then we have

$$(35) \qquad (f(z))^{-} = (2\pi i)^{-1}\int_{C_1} K_B(w,\bar{z})\mathfrak{B}(g(w))\,dw,$$

where C_1 is any simple closed curve lying in the interior of B and containing $D(g)$ in its interior.

Proof. With $p_n(z) = \sum_{k=0}^{n} a_{nk}z^k$, we have by (9)

$$(36) \qquad g(d)p_n(z) = \sum_{k=0}^{n} a_{nk}g^{(k)}(0) = \sum_{k=0}^{n} a_{nk}(w^k, f) = (p_n, f).$$

Hence the Fourier coefficients of f are $(g(d)p_n(z))^{-}$, and (32) follows. Equation (33) follows from (32) by Theorem 3. To arrive at (34), we notice that by Cauchy's theorem and (24),

$$(37) \qquad (2\pi i)^{-1}\int_{C} g(z)\mathfrak{B}(p_n(z))\,dz = \sum_{k=0}^{n} a_{nk}g^{(k)}(0) = (p_n, f).$$

To establish (35), we proceed as follows. Let C designate a circle centered at the origin and of radius sufficiently large so that $D(g)$ is contained in its interior. If $g(z) = \sum_{n=0}^{\infty}\gamma_n z^n$, then the function $\mathfrak{B}(g(z)) = \sum_{n=0}^{\infty}\gamma_n n!/z^{n+1}$ is regular on C_1, while its Laurent series converges absolutely and uniformly on C. Thus we have

$$\begin{aligned}(2\pi i)^{-1}\int_{C_1} p_n(z)\mathfrak{B}(g(z))\,dz &= (2\pi i)^{-1}\int_{C} p_n(z)\mathfrak{B}(g(z))\,dz \\ &= (2\pi i)^{-1}\int_{C} p_n(z)\sum_{k=0}^{\infty}\gamma_k k!/z^{k+1}dz = \sum_{k=0}^{n} a_{nk}\gamma_k k! \\ &= \sum_{k=0}^{n} a_{nk}g^{(k)}(0) = (2\pi i)^{-1}\int_{C} g(z)\mathfrak{B}(p_n(z))\,dz.\end{aligned}$$

The last equality follows by (37). By (34), therefore,

$$(f(z))^{-} = \sum_{n=0}^{\infty}(p_n(z))^{-}(2\pi i)^{-1}\int_{C_1} p_n(z)\mathfrak{B}(g(z))\,dz.$$

Since $K_B(w,\bar{z})$ is given by (3), and for fixed z converges uniformly for w in any closed subset of B, we arrive at (35) easily.

COROLLARY. *Let $g = \mathfrak{L}_B(f)$. Then for all $h(z)\,\varepsilon\, L^2(B)$,*

$$(38) \qquad g(d)_B h(z) = (h, f).$$

Proof. This follows directly from Theorems 3 and 6.

COROLLARY. *If* $g(z) \equiv 0$, *then* $f(z) \equiv 0$.

Two functions whose $\mathfrak{L}_B$ transforms are identical are therefore themselves identical. It should be observed that this result has been derived on the assumption that B is a domain for which the powers of z form a complete system. It will be shown in a later paragraph that there exist bounded simply-connected domains for which the $\mathfrak{L}_B$ transformation is not one-to-one.

We next record a number of identities relating the orthonormal polynomials and their $\mathfrak{L}_B$ transforms $q_n(z)$.

COROLLARY. *The sets* $\{p_n(z)\}$ *and* $\{q_n(z)\}$ *are biorthonormal in the sense that*

$$q_m(d)p_n(z) = p_n(d)q_m(z) = \delta_{mn} \qquad (m, n = 0, 1, \cdots). \tag{39}$$

This results from the first corollary above and the orthonormality of the p_n's. The biorthonormality relations (39) may also be expressed in the following two ways. Let A and B be the ∞ by ∞ matrices of coefficients

$$A = (a_{nk}), \quad B = (b_{nk}), \tag{40}$$

with a_{nk} given by (4) and b_{nk} by (31). Then

$$AB = BA = I. \tag{41}$$

Alternately, the following identity holds for w arbitrary and uniformly for z lying in a closed subdomain of B.

$$e^{zw} = \sum_{n=0}^{\infty} q_n(w)p_n(z). \tag{42}$$

In section 3 we proved that if $g(z) \varepsilon E^2(B)$, then $D(g) \subseteq \hat{B}$. In addition, we gave an example showing that the boundaries of $D(g)$ and of $\hat{B}$ may have points in common. In the reverse direction, we shall show in Theorem 7 that if $D(g)$ is contained in the interior of B, then $g(z)$ is surely a member of $E^2(B)$. But if we know only that $D(g) \subseteq \bar{B}$, it may not follow that $g \varepsilon E^2(B)$.

As an example of this behavior, let B be selected as the unit circle. For $g(z)$, choose the entire function $g(z) = (e^z - 1)/z$. It is easily verified that $\mathfrak{B}(g) = \log(1 - z^{-1})$. Therefore $D(g)$ has a point which lies on the boundary of B, viz. $z = 1$. Again, we have $p_n(z) = (\pi)^{-\frac{1}{2}}(n+1)^{\frac{1}{2}}z^n$, and assuming that $g = \mathfrak{L}_B(f)$, where $f = \sum_{n=0}^{\infty} \alpha_n p_n$, we find that $\alpha_n = (\pi(n+1))^{-\frac{1}{2}}$. Thus $\| f \| = \infty$, and g is the transform of no $f \varepsilon L^2(B)$.

THEOREM 7. *Let $g(z)$ be an entire function of exponential type and such that $D(g) \subset B$. Then $g(z) \varepsilon E^2(B)$.*

Proof. Let C_1 be a simple closed rectifiable contour lying in B and containing $D(g)$ in its interior. Construct the function

$$(43)\qquad f(z) = \sum_{n=0}^{\infty} \bar{k}_n p_n(z),$$

where

$$(44)\qquad k_n = (2\pi i)^{-1} \int_{C_1} p_n(z) \mathfrak{B}(g(z))\, dz \qquad (n = 0, 1, \cdots).$$

Applying the Schwarz inequality to (44), we obtain

$$(2\pi)^2 \mid k_n \mid^2 \leqq \mathrm{l}(C_1) \max_{z \varepsilon C_1} \mid \mathfrak{B}(g(z)) \mid^2 \int_{C_1} \mid p_n(z) \mid^2 ds,$$

so that

$$(2\pi)^2 \sum_{n=0}^{\infty} \mid k_n \mid^2 \leqq \mathrm{l}(C_1) \max_{z \varepsilon C_1} \mid \mathfrak{B}(g(z)) \mid^2 \int_{C_1} K_B(z, \bar{z})\, ds < \infty,$$

inasmuch as C_1 is interior to B. Therefore $f \varepsilon L^2(B)$, and we have

$$g(z) = (2\pi i)^{-1} \int_{C_1} e^{zw} \mathfrak{B}(g(w))\, dw = (2\pi i)^{-1} \int_{C_1} \sum_{n=0}^{\infty} q_n(z) p_n(w) \mathfrak{B}(g(w))\, dw$$
$$= \sum_{n=0}^{\infty} q_n(z) k_n = \mathfrak{L}_B(f(z))$$

by (25), (42), (43), and (30) successively.

COROLLARY. *Let B_1 and B_2 be two domains such that $\hat{B}_1 \subset B_2$. Then $E^2(B_1) \subset E^2(B_2)$. In particular, the conclusion holds if $B_1 \subset B_2$ and at least one of the two domains is convex.*

Proof. Let $g(z) \varepsilon E^2(B_1)$. Then $D(g) \subseteq \hat{B}_1 \subset B_2$ by Theorem 5 and our hypothesis. Therefore, by the present theorem, $g(z) \varepsilon E^2(B_2)$.

Under the hypotheses of Theorem 7, an alternate expression for $g(d)_B f(z)$, $f \varepsilon L^2(B)$, may be found. From (38) and (35) we obtain

$$(45)\qquad g(d)_B f(z) = \iint_B f(z)\, (2\pi i)^{-1} \int_{C_1} K_B(w, \bar{z}) \mathfrak{B}(g(w))\, dw\, dxdy,$$

where C_1 is a simple closed curve lying in the interior of B and containing $D(g)$ in its interior. Thus we have

$$(46)\qquad g(d)_B f(z) = (2\pi i)^{-1} \int_{C_1} \mathfrak{B}(g(w))\, dw \iint_B f(z) K_B(w, \bar{z})\, dxdy$$
$$= (2\pi i)^{-1} \int_{C_1} f(w) \mathfrak{B}(g(w))\, dw.$$

If it is further assumed that B is a star domain with respect to the origin, then by the well-known theory of Mittag-Leffler summability of Taylor series we have

$$(47)\qquad f(w)=\lim_{\delta\to\infty}\sum_{n=0}^{\infty}f^{(n)}(0)w^n/n!\Gamma(1+(n-1)\delta),$$

uniformly on every closed subset of B. With (46) this yields

$$(48)\qquad g(d)_B f=\lim_{\delta\to 0}\sum_{n=0}^{\infty}(2\pi i)^{-1}\int_{C_1}w^n\mathfrak{B}(g(w))\,dw\,f^{(n)}(0)/n!\Gamma(1+(n-1)\delta)$$
$$=\lim_{\delta\to 0}\sum_{n=0}^{\infty}a_n f^{(n)}(0)/\Gamma(1+(n-1)\delta),$$

where we have assumed that $g(z)=\sum_{n=0}^{\infty}a_n z^n$. This may be written

$$(49)\qquad g(d)_B f(z)=(M)\sum_{n=0}^{\infty}a_n f^{(n)}(0),$$

the symbol (M) indicating a process of Mittag-Leffler summation. If it is known from independent considerations that $\sum_{n=0}^{\infty}a_n f^{(n)}(0)$ converges, then since (M) is a regular method of summation, we have

$$(50)\qquad g(d)_B f(z)=\sum_{n=0}^{\infty}a_n f^{(n)}(0)=g(d)f(z).$$

5. Completeness of the set $\{e^{\lambda_n z}\}$. In the present section, we employ the inversion formula just derived plus the well-known device of Szász [10] to obtain conditions on the distribution of the complex values λ_n under which the set of functions $\{e^{\lambda_n z}\}$ $(n=0,1,\cdots)$ will be complete for $L^2(B)$. By this device, the completeness is referred to the possible distribution of zeros of the entire functions of class $E^2(B)$.

THEOREM 8. *Let $\lambda_0,\lambda_1,\cdots$ be a sequence of distinct complex numbers with $\lambda_n\to\infty$. Let $n(x)$ denote the number of λ's for which $|\lambda_n|\leqq x$. Let $c(\hat{B})$ designate the circumference of the convex hull of B. If*

$$(51)\qquad \limsup_{r\to\infty} r^{-1}\int_1^r (n(x)/x)\,dx > c(\hat{B})/2\pi,$$

then the set $\{e^{\lambda_n z}\}$ will be complete in $L^2(B)$.

Proof. Suppose that the sequence of λ's satisfies (51) but that the set is not complete. Then there will exist an $f\in L^2(B)$ not identically zero, such that $(e^{\lambda_n z}, f(z))=0$ $(n=0,1,\cdots)$. Consider now the function $g=\mathfrak{L}_B(f)$.

The entire function $g(z)$ is not identically zero by the second corollary to Theorem 6. The points $z=\lambda_n$ form a subset of its zeros. If, therefore, $n^*(x)$ denotes the number of zeros of g whose moduli do not exceed x, we shall have $n(x) \leqq n^*(x)$. By Jensen's Theorem (see, e. g., Titchmarsh (11) we have

$$(52)\quad \int_1^r (n(x)/x)\,dx \leqq \int_1^r (n^*(x)/x)\,dx < (2\pi)^{-1}\int_0^{2\pi} \log|g(re^{i\theta})|\,d\theta + c,$$

where c is a fixed constant which depends solely upon g. Now from Theorem 5 we have $\log|g(re^{i\theta})| \leqq rk(-\theta) + \log\|f\| + \frac{1}{2}\log \text{Area}(B)$, so that

$$\limsup_{r\to\infty} r^{-1}\int_1^r (n(x)/x)\,dx \leqq (2\pi)^{-1}\int_0^{2\pi} k(-\theta)\,d\theta = c(\hat{B})/2\pi$$

(on this last equality see [8]). This contradicts (51).

It is possible to show that for convex domains possessing a center of symmetry, the constant $c(\hat{B})/2\pi$ is the best possible. This fact falls out as a corollary to a more general theorem which we shall now prove.

THEOREM 9. *Let the domain B contain in its interior a convex polygon P which posesses a center of symmetry. Then there exist systems $\{e^{\lambda_n z}\}$ with*

$$(53)\qquad \lim_{r\to\infty} r^{-1}\int_1^r (n(x)/x)\,dx = c(P)/2\pi$$

which are not complete in $L^2(B)$.

Proof. We assume that $z=0$ is the center of symmetry of P. Select a vertex of P, denote it by z_1, and, proceeding counterclockwise, let the remaining vertices be denoted by $z_2, z_3, \cdots, z_n, -z_1, -z_2, \cdots, -z_n$. Set

$$(54)\qquad \begin{aligned} &w_1 = \tfrac{1}{2}(z_2 - z_1),\quad w_2 = \tfrac{1}{2}(z_3 - z_2), \cdots, w_{n-1} = \tfrac{1}{2}(z_n - z_{n-1}),\\ &w_n = \tfrac{1}{2}(-z_1 - z_n).\end{aligned}$$

Then the circumference of P is $c(P) = 4(|w_1| + |w_2| + \cdots + |w_n|)$. Consider now the function

$$(55)\qquad \begin{aligned}\phi(z) &= \prod_{j=1}^n \sin(w_j z/i) = (-i/2)^n \prod_{j=1}^n (e^{w_j z} - e^{-w_j z})\\ &= (-i/2)^n \sum \pm e^{\sum_{j=1}^n \pm w_j z}\end{aligned}$$

where the first sum is to be extended over the 2^n possible combinations of $\pm$ in the second sum. It follows from (24) that the singularities of $\mathfrak{B}(\phi(z))$

are located at the 2^n points $\sum_{j=1}^{n} \pm w_j$. Now it may be easily verified that the $2n$ vertices of P are located at points $\sum_{j=1}^{n} \pm w_j$ for some $2n$ selections of $\pm$, while the remaining $2^n - 2n$ selections give rise to points which are interior to P. Thus the conjugate indicator diagram of ϕ is precisely $P: D(\phi) = P$. Since P is interior to B, by Theorem 7, $\phi(z) = \mathfrak{L}_B(\psi)$ for some $\psi \varepsilon L^2(B)$. If $\lambda = m\pi i/w_j$ $(m = 0, 1, 2, \cdots; j = 1, 2, \cdots, n)$, then we have by (55), $(e^{\lambda z}, \psi) = 0$. Let λ_k denote these points arranged in order of non-decreasing absolute value. For this set it is clear that

$$(56) \qquad n(x) = (2x/\pi)(|w_1| + |w_2| + \cdots + |w_n|) + O(1), \qquad x \to \infty.$$

Hence for the function ψ we have

$$\lim_{r\to\infty} (r^{-1}) \int_1^r (n(x)/x)\,dx = (2/\pi)(|w_1| + \cdots + |w_n|) = c(P)/2\pi.$$

COROLLARY. *If B is convex and possesses a center of symmetry, then there exist systems $\{e^{\lambda_n z}\}$ for which*

$$(57) \qquad \lim_{r\to\infty} (r^{-1}) \int_1^r (n(x)/x)\,dx = c(B)/2\pi - \epsilon,$$

and which are not complete in $L^2(B)$.

Proof. In this case we may find a polygon possessing the same center of symmetry, and lying in the interior of B, whose perimeter differs from that of B by arbitrarily small amounts.

Throughout the present work the assumption has been made that the domain B is such that the powers of z are complete for $L^2(B)$. Under this assumption, the $\mathfrak{L}_B$ transform has been shown to be one-to-one. However this transform may be defined for any bounded domain, and the order conditions of Theorem 5 will be preserved. But it is possible to show that in general the transform will not produce a one-to-one mapping of $L^2(B)$ on $E^2(B)$. To show this, suppose that the domain B, assumed to satisfy our usual conditions, is slit. That is, let a Jordan arc issuing from a boundary point of B and penetrating into its interior be added as a part of its boundary. Denote the resulting slit domain by B_s. The following theorem now holds.

THEOREM 10. *There exists a function $f(z) \varepsilon L^2(B_s)$ which is not identically zero, and is such that*

$$(58) \qquad \mathfrak{L}_{B_s}(f) = 0.$$

Prooof. Assume that the conclusion is false and that the only function for which (58) holds is the null function. If a sequence of constants λ_n is selected so that condition (51) is satisfied, then the proof of Theorem 8 will go through word for word, and we will have proved that the set $\{e^{\lambda_n z}\}$ is complete for the class $L^2(B_s)$. Now this is impossible, for if we orthonormalize the set $\{e^{\lambda_n z}\}$ with respect to B_s, the resulting set of exponential polynomials will be complete and orthonormal for B itself, inasmuch as the slit does not affect the values of the area integrals which occur in the orthonormalization process. If, then, $f(z) \,\varepsilon\, L^2(B_s)$, it would have an expansion in orthogonal exponential polynomials converging in the interior of B, and we would have every function of $L^2(B_s)$ regular in B. This is a contradiction.

6. Application to a coefficient problem in the theory of entire functions of exponential type. Let $f(z)$ be an entire function of exponential type. As we have already mentioned, the indicator of f is essentially the supporting function of the convex domain $D(f): h_f(\theta) = k_{D(f)}(-\theta)$, $0 \leqq \theta \leqq 2\pi$. Conversely, it is known that for any given convex domain C there are entire functions f such that $D(f) = \bar{C}$. We now pose the following problem. Let $L(z) = \sum_{n=0}^{\infty} \beta_n z^n$. What are necessary and sufficient conditions on the sequence of coefficients $\{\beta_n\}$ in order that $L(z)$ have a preassigned indicator, or what is equivalent, in order that $D(L(z))$ be a given convex domain. Such a condition is conveniently given in terms of the norm of the operator $L(d)$.

THEOREM 11. *Let the domain B be convex and $k_B(\theta)$ denote its supporting function. In order that the indicator $h(\theta)$ of $L(z)$ satisfy*

$$(59) \qquad h(\theta) = k_B(-\theta), \qquad 0 \leqq \theta \leqq 2\pi,$$

it is necessary and sufficient that

$$(60) \qquad \| L(d) \|_{B^*} < \infty \qquad \text{for all } B^* \text{ with } \bar{B} \subset B^*,$$

and

$$(61) \qquad \| L(d) \|_{B^*} = \infty \qquad \text{for all } B^* \text{ with } B \cap \operatorname{Ext} B^* \neq 0.$$

Proof. Let (60) and (61) hold. By (60), $L(z) \,\varepsilon\, E^2(B^*)$ for all B^* containing B in its interior. Select a sequence of convex domains B^*_n $(n = 1, 2, \cdots)$, such that $B^*_n \supseteq B^*_{n+1}$, and which converge to B. If $k_{B^*_n}(\theta)$ are the supporting functions of B^*_n, then we have $\lim_{n\to\infty} k_{B^*_n}(\theta) = k_B(\theta)$, $0 \leqq \theta \leqq 2\pi$, so that since $h(\theta) \leqq k_{B^*_n}(-\theta)$, we have $h(\theta) \leqq k_B(-\theta)$ $(0 \leqq \theta \leqq 2\pi)$. If now $h(\theta) < k_B(-\theta)$ for some θ, then by the continuity

of $h(\theta)$, the inequality would hold over a range of values, and therefore B would have interior points which are exterior to $D(L(z))$. It would then be possible to find a domain B' containing $D(L(z))$ in its interior but such that B has points exterior to B'. Since $D(L(z))$ is interior to B', by Theorem 7, $L(z) \varepsilon E^2(B')$. This is impossible by (61). Therefore $h(\theta) = k_B(-\theta)$, $0 \leqq \theta \leqq 2\pi$.

Conversely, suppose that (59) holds. Then $D(L(z)) = \bar{B}$. Let B^* contain B in its interior; then by Theorem 7, $L(z) \varepsilon E^2(B^*)$, so that (60) must hold. Suppose that B contains an interior point exterior to a domain B^*, and that (61) does not hold. Then $L(z) \varepsilon E^2(B^*)$. From Theorem 5, $D(L(z))$ must lie in $\bar{B}^*$ and we have a contradiction.

If the orthonormal polynomials of a domain B^* are given by

$$p^*_n(z) = \sum_{k=0}^{n} a^*_{nk} z^k \qquad (n = 0, 1, \cdots),$$

then conditions (60) and (61) become

$$\sum_{n=0}^{\infty} \Big| \sum_{k=0}^{n} a^*_{nk} \beta_k k! \Big|^2 < \infty \text{ or } = \infty. \tag{62}$$

Since a^*_{nk} depend solely upon B^*, we see that (60), (61) are, in reality, conditions to be satisfied by the coefficients of $L(z)$.

As a special case, we may obtain the familiar Lindelöf relationship between the coefficients of the series development of an entire function of exponential type and the type. Let $g(z) = \sum_{n=0}^{\infty} \beta_n z^n$ be an entire function of exponential type σ. A necessary and sufficient condition for this is that $D(g)$ be contained in the circle $|z| \leqq \sigma$, and that at least one boundary point of $D(g)$ lie on $|z| = \sigma$. If C_r designates the circle $|z| = r$, then it is necessary and sufficient that

$$\| g(d) \|_{C_r} < \infty \quad \text{for} \quad r > \sigma, \tag{63}$$

$$\| g(d) \|_{C_r} = \infty \quad \text{for} \quad r < \sigma. \tag{64}$$

Referring to (13), $\| g(d) \|^2_{C_r} = \sum_{n=0}^{\infty} |\beta_n|^2 n!(n+1)! r^{-2n}$, while the radius of convergence of this series in r is

$$\sigma = \limsup_{m \to \infty} (m/e) |\beta_m|^{1/m}. \tag{65}$$

Theorem 11 can therefore be regarded as a generalization of Lindelöf's Theorem.

HARVARD UNIVERSITY.

BIBLIOGRAPHY.

[1] N. Aronszajn, "Theory of reproducing kernels," *Transactions of the American Mathematical Society*, vol. 68 (1950), pp. 337-404.

[2] S. Bergman, "The kernel function and conformal mapping," *Mathematical Surveys*, No. 5, New York, 1950.

[3] R. P. Boas, Jr., "Entire functions of exponential type," *Bulletin of the American Mathematical Society*, vol. 48 (1942), pp. 839-849.

[4] R. C. Buck, "Interpolation and uniqueness of entire functions," *Transactions of the American Mathematical Society*, vol. 64 (1948), pp. 283-298.

[5] O. J. Farrell, "On the approximation to an analytic function by polynomials," *Bulletin of the American Mathematical Society*, vol. 40 (1934), pp. 908-914.

[6] E. Landau, "Über einen Konvergenzsatz," *Nachrichten der Königlichen Gesellschaft der Wissenschaften zu Göttingen* (1907), pp. 25-27.

[7] H. Muggli, "Differentialgleichungen unendlichhoher Ordnung mit konstanten Koeffizienten," *Comentarii Mathematici Helvetici*, vol. 11 (1938), pp. 151-179.

[8] G. Pólya, "Untersuchungen über Lücken und Singularitaten von Potenzreihen," *Mathematische Zeitschrift*, vol. 29 (1929), pp. 549-640.

[9] M. H. Stone, "Linear transformations in Hilbert space," *American Mathematical Society Colloquium Publication*, vol. XV, New York, 1932.

[10] O. Szász, "über die Approximation stetiger Funktionen durch lineare Aggregate von Potenzen," *Mathematische Annalen*, vol. 77 (1916), pp. 482-496.

[11] E. C. Titchmarsh, *The Theory of Functions*, 2nd ed., London, 1939.

UNIQUENESS THEOREMS FOR TWO FREE BOUNDARY PROBLEMS.*

By James B. Serrin, Jr.

An important new approach to the Helmholtz problem of plane wakes, based upon general results in the theory of conformal representation, has been developed by Lavrentieff [1]. In his paper Lavrentieff considered the general existence and uniqueness for a certain class of obstacles; however, the essential simplicity and elegance of his method in the case of the uniqueness problem is obscured by his recourse to results of the complicated existence theory as well as by the difficulty one experiences in understanding his proof.[1]

It is the purpose of this paper to present an adequate and simple account of this uniqueness theory and to extend its validity. In Part I we develop the function theoretic basis for the theory. Theorem 1 is Lavrentieff's basic result, but the proof we give is new; on the other hand, Theorem 2 was neither stated nor proved by Lavrentieff, although he apparently made use of the result. In Part II the uniqueness theory proper is presented. In its exposition we have used, outside of Theorems 1 and 2, only one well known property of a flow solving the plane wake problem. The class of obstacles we consider includes as a special case the class considered by Lavrentieff, while it neither includes the class of obstacles discussed by Leray [2] nor is it included in that class.

Finally, in § 6 we show that flows solving the so-called finite cavity problem exist, while in § 7 we indicate certain generalizations.

Part I.

1. A corollary of Julia's theorem. Julia's theorem may be stated in the following form:[2] *Let* $f(\zeta)$ *be regular and* $|f(\zeta)| < 1$ *in* $|\zeta| < 1$. *Suppose*

* Received February 26, 1951. Prepared under Navy Contract N6onr-180, Task Order V, with Indiana University.

[1] The present author states frankly that he is unable to follow many of the arguments. Also see Weinstein, *Proc. Symposia on Applied Math.*, vol. 1, Am. Math. Soc., 1949, pp. 1-12.

[2] Bieberbach, Lehrbuch der Funktionentheorie, Chelsea, 1945, vol. 2, pp. 112-121.

that $f(1) = 1$ *and* $\lim f'(\zeta) = \alpha$, *where* $\zeta \to 1$. *Then* α *is real and positive, and*

$$(1) \qquad |1 - f(\zeta)|^2/(1 - |f(\zeta)|^2) \leq \alpha\, |1 - \zeta|^2/(1 - |\zeta|^2).$$

The equality holds if and only if $f(\zeta)$ *is linear.*

COROLLARY. *Let* $f(\zeta)$ *satisfy the hypotheses of Julia's theorem. In addition, suppose that* $f(-1) = -1$ *and* $\lim f'(\zeta) = \beta$, *where* $\zeta \to -1$. *Then* β *is real and positive, and* $\alpha\beta \geq 1$. *The equality holds if and only if*

$$(2) \qquad f(\zeta) = (\zeta + a)/(a\zeta + 1)$$

where a *is a real number,* $|a| < 1$.

Proof. The hypotheses of Julia's theorem are satisfied by the function $g(\zeta) = -f(-\zeta)$; in fact $\lim_{\zeta \to 1} g'(\zeta) = \lim_{\zeta \to 1} f'(-\zeta) = \beta$. Hence β is real and positive, and

$$(3) \qquad |1 + f(-\zeta)|^2/(1 - |f(-\zeta)|^2) \leq \beta\, |1 - \zeta|^2/(1 - |\zeta|^2).$$

Putting $\zeta = 0$ in (1) and (3), and multiplying the resulting inequalities together, we obtain

$$\alpha\beta \geq |1 - f(0)^2|^2/(1 - |f(0)|^2)^2 \geq 1.$$

Obviously, the equality can hold only if $f(\zeta)$ is linear. However, the only linear function satisfying the hypotheses is (2). For this function, $\alpha\beta = 1$, which completes the proof.

2. Applications to hydrodynamics. By a *flow* we shall mean a steady irrotational plane motion of an ideal incompressible fluid. We suppose that the motion occurs in a closed simply connected region R (which may be many sheeted) over the complex z-plane. The fluid may not penetrate the boundary of R (that is the boundary consists of streamlines), although a finite number of sources and sinks are allowed on the boundary. Let $w(z) = \phi + i\psi$ be the usual complex velocity potential; then $w(z)$ is regular in the interior of R and maps the boundary of R into straight arcs $\psi =$ const. of the w-plane.

Without exception, each flow we consider will occupy an infinite region and have a well defined velocity at infinity of fixed magnitude U.[3] A flow of this type will be called an *admissible flow* if $w(z)$ maps R onto the upper half plane $\psi \geq 0$. If in addition the inverse of $w(z)$ is single valued and

[3] By this we mean that the complex velocity $w'(z) = u + iv$ approaches a limit at infinity, and $|w'(\infty)| = U$.

regular for $\psi > 0$, then the flow will be called a *simple flow*. The speed at a point P in R of a flow in R will be denoted by $V(P, R)$; i. e.

$$V(P, R) \equiv |\, w'(z_P)\,|.^4$$

Since only one flow will be considered corresponding to a given region R, this notation will cause no confusion.

Let R_1 and R_2 be two regions of the type described above. If each point P of the surface R_2 may be set in correspondence with a point P' of R_1 such that $z_P = z_{P'}$, and if this correspondence is continuous for P on R_2, then we say that R_1 contains R_2. We shall often be dealing with several flows simultaneously; although these flows are considered as occurring in the same plane (z-plane), they are independent of one another. Our results will depend on the following theorems.

THEOREM 1. *Let R_1 and R_2 be two closed simply connected regions over the complex z-plane, such that R_1 contains R_2. Suppose that there exists an admissible flow in R_1 and a simple flow in R_2. If the streamlines $\psi = 0$ have a common point M, then $V(M, R_1) \geq V(M, R_2)$, the equality holding if and only if the flows are identical, or $V(M, R_1) = 0$.*

Proof. Let $w_1(z)$ and $w_2(z)$ denote the complex velocity potentials of the flows in R_1 and R_2 respectively. Without loss of generality we can take $w_1(z_M) = w_2(z_M) = 0$.

Define

$$\zeta(w) = \{-(i-1)w - (i+1)\}/\{(i-1)w - (i+1)\}.$$

This function maps the upper half plane, which we shall denote by Δ, onto the unit circle C such that $\zeta(0) = 1$, $\zeta(\infty) = -1$. If the unit circle C is mapped upon the upper half plane Δ by the inverse of $\zeta(w)$, Δ onto R_2 by the single valued regular function which is the inverse of $w_2(z)$, R_2 into a subregion of Δ by $w_1(z)$, and this subregion into C by $\zeta(w)$, we achieve a conformal mapping $f(\zeta)$ of the unit circle into itself. Clearly $f(1) = 1$, $f(-1) = -1$,

$$f'(\zeta) = \frac{\zeta'[w_1(z)] \cdot w_1'(z)}{\zeta'[w_2(z)] \cdot w_2'(z)} = \frac{w_1'(z)}{w_2'(z)} \left[\frac{(i-1)w_2(z) - (i+1)}{(i-1)w_1(z) - (i+1)}\right]^2.$$

---

[4] The complex number corresponding to a point P in the flow plane will be denoted by z_P. If P is on the boundary of R, $w'(z_P)$ does not necessarily exist. However, by writing the symbol $V(P, R)$ we shall imply that the derivative in question does exist.

Hence, recalling that α and β are real,

$$\alpha = \lim_{\zeta\to 1} |f'(\zeta)| = \lim_{z\to z_M} \left|\frac{w_1'(z)}{w_2'(z)}\right| \left|\frac{(i-1)w_2(z)-(i+1)}{(i-1)w_1(z)-(i+1)}\right|^2$$
$$= V(M,R_1)/V(M,R_2);$$
$$\beta = \lim_{\zeta\to -1} |f'(\zeta)| = \lim_{z\to\infty} | w_2(z)/w_1(z)|^2.$$

To evaluate the limit, we fix $\epsilon > 0$. For z sufficiently large, $| w_1'(z) - U | < \epsilon$, $| w_2'(z) - U | < \epsilon$. Using these inequalities, we obtain for sufficiently large z, z_A:

$$\frac{(U-\epsilon)|z-z_A| - |w_2(z_A)|}{(U+\epsilon)|z-z_A| + |w_1(z_A)|} \leq \left|\frac{w_2(z)}{w_1(z)}\right| \leq \frac{(U+\epsilon)|z-z_A| + |w_2(z_A)|}{(U-\epsilon)|z-z_A| - |w_1(z_A)|}.$$

Hence $(U-\epsilon)/(U+\epsilon) \leq \lim_{z\to\infty} | w_2(z)/w_1(z)| \leq (U+\epsilon)/(U-\epsilon)$. Since ϵ can be made arbitrarily small, the limit is 1. Thus, $\beta = 1$.

Applying the corollary, we get $\alpha\beta = V(M,R_1)/V(M,R_2) \geq 1$. The equality holds if and only if $f(\zeta) = (\zeta + a)/(a\zeta + 1)$; that is, if and only if the flows in R_1 and R_2 are identical.

The following lemma will be useful in proving Theorems 2 and 3.

LEMMA 1. *Let R_1 and R_2 be two regions whose boundaries have a common arc MN, and which lie on the same side of MN. Suppose $w_1(z)$ and $w_2(z)$, respectively, map R_1 and R_2 onto the upper half plane. If P is an interior point of MN and* $\lim | w_1'(z)|$, *where $z \to z_P$, exists, then* $\lim | w_2'(z)|$ *exists, where $z \to z_P$.*

The simple proof, which we omit, depends on the Schwarz reflection principle.

THEOREM 2. *Let R_1 and R_2 be two simply connected closed regions over the complex-z-plane. Suppose that there are admissable flows in R_1 and R_2, and let S_1 and S_2 denote the respective streamlines $\psi = 0$. Let S_1 and S_2 have an arc MN in common such that the direction of each flow on MN is from M to N. Furthermore let an arc QM of S_1 and an arc NQ of S_2 bound, together with MN, a simply connected region $MNQ \equiv R_3$ which is contained in both R_1 and R_2. Then the inequality*

$$V(M,R_2) < V(N,R_2), \tag{4}$$

implies that $V(M,R_1) < V(N,R_1)$.

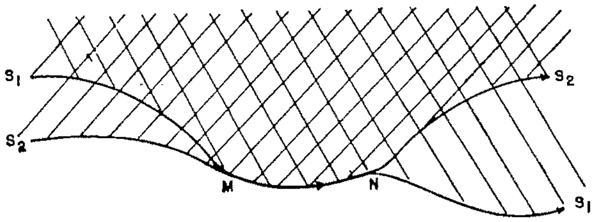


FIGURE 1

Note. Figure 1 shows a particularly simple case of the theorem when Q is the point at infinity and the flows in both R_1 and R_2 are simple. It is convenient to think of this situation rather than the more general one during the course of the proof, although the argument is identical in any case.[5]

Proof. Let $w_2(z)$ be the complex potential of the admissable flow in R_2. We may take $w_2(z_M) = 0$. Let $w_3(z)$ map R_3 conformally onto the upper half plane Δ such that $w_3(z_M) = 0$, $w_3(z_N) = w_2(z_N)$, $w_3(z_Q) = \infty$. By virtue of Lemma 1, $\lim_{z \to z_M} |w_3'(z)|$ and $\lim_{z \to z_N} |w_3'(z)|$ exist; we shall denote these limits by $V(M, R_3)$ and $V(N, R_3)$. The proof rests on two auxiliary inequalities.

First we shall show that

$$V(N, R_2)/V(N, R_3) \leq 1. \tag{5}$$

Let $\zeta(w)$ map Δ onto the upper half unit circle Γ such that $\zeta(0) = -1$, $\zeta[w_2(z_N)] = 0$, $\zeta(\infty) = +1$. The image of the positive real axis of the w-plane is then the segment $(-1, +1)$ of the real axis in the ζ-plane.

If Γ is mapped onto Δ by the inverse of $\zeta(w)$, Δ onto R_3 by the inverse of $w_3(z)$, R_3 into a subregion of Δ by $w_2(z)$, and this subregion into Γ by $\zeta(w)$, we achieve a mapping $f(\zeta)$ of Γ into itself. Now $f(0) = 0$, and $f(\zeta)$ is real and continuous on the real axis. Consequently, $f(\zeta)$ may be analytically continued, by reflection, throughout the entire unit circle. Then by Schwarz's Lemma, $f'(0) \leq 1$. But

$$f'(0) = \{\zeta'[w_2(z_N)]w_2'(z_N)\}/\{\zeta'[w_3(z_N)]w_3'(z_N)\} = V(N, R_2)/V(N, R_3).$$

Secondly, we shall show that

$$\{V(M, R_2)/V(M, R_3)\}\{V(N, R_2)/V(N, R_3)\} \geq 1. \tag{6}$$

[5] Theorem 2, as well as Theorem 1, is stated in more generality than is necessary for the purposes of this paper; in particular, the flows we consider in applications will all be simple. However, the increase in generality does not alter the proof, and in a later paper [3] it will be of the utmost importance.

It may be remarked here that if equality holds in (4), then equality holds in the conclusion if and only if the flows are identical.

Using a series of conformal mappings very similar to that already exhibited, we may define a function $g(\zeta)$ such that $|g(\zeta)| < 1$ for $|\zeta| < 1$, and $g(1) = 1$, $g(-1) = -1$, finally $g'(1) = V(M, R_2)/V(M, R_3)$, $g'(-1) = V(N, R_2)/V(N, R_3)$. The required result follows from our corollary.

Combining inequalities (5) and (6) we obtain

$$(7) \qquad V(N, R_2)/V(N, R_3) \leq 1 \leq V(N, R_3)/V(N, R_2) \leq V(M, R_2)/V(M, R_3).$$

Hence, using (7) and (4),

$$(8) \qquad V(M, R_3) \leq \{V(M, R_2)/V(N, R_2)\} V(N, R_3) < V(N, R_3).$$

Now a duplication of the preceding argument, this time with respect to the regions R_3 and R_1, leads to

$$(7') \qquad V(M, R_1)/V(M, R_3) \leq V(N, R_1)/V(N, R_3);$$

(it is sufficient to interchange the roles of M and N, and to replace R_2 by R_1). From (7′) and (8) we obtain

$$V(M, R_1) \leq \{V(M, R_3)/V(N, R_3)\} V(N, R_1) < V(N, R_1),$$

which completes the proof.

The following theorem is essentially due to Lavrentieff. Although no use will be made of its result, it is an interesting corollary of the preceding theorems. Let R be a region occupied by a simple flow, and let S be the streamline $\psi = 0$. We may assume without loss of generality that S has zero tangent at infinity and that R is "above" S. Let S contain an arc M^*PN^* having the following properties:

1. If S is given parametrically by $z = z(\phi)$, then the points $-\infty$, M^*, P, N^*, $+\infty$ correspond in that order to increasing ϕ;

2. M^*PN^* has a continuous tangent;

3. P is an inclination point of S, and the inclination of S is monotone on M^*P and PN^*;

4. The straight line L tangent to S at P does not intersect S other than at P, and is not horizontal.

5. The speed $V(Q, R)$ exists at all interior points Q of M^*PN^*.

A direction is assigned to L to correspond to the direction of increasing ϕ at P; Figure 2.

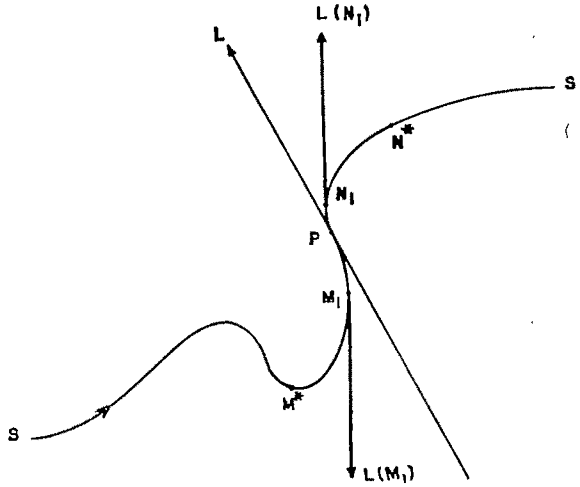


FIGURE 2

THEOREM 3. *Under hypotheses* 1-5, *there is an arc* MPN *of* M^*PN^* *such that, if* L *is directed into* R, $V(N,R) > V(M,R)$, *or if* L *is directed out of* R, $V(N,R) < V(M,R)$.

Proof. We note here that no attempt has been made to obtain a minimal list of hypotheses for the validity of the theorem, or even to state the present hypotheses in the shortest possible form. It is sufficient to consider only the case that L is directed into R.

Let N_0 be any point of PN^*, and denote by $L(N_0)$ the ray tangent to S at N_0, directed in the sense of increasing ϕ at N_0. Similarly let M_0 be any point of M^*P, and denote by $L(M_0)$ the ray tangent to S at M_0, directed in the sense of decreasing ϕ at M_0.

In virtue of hypotheses 1-4 there exists an arc M_1PN_1 of M^*PN^* such that $L(N_1)$ is entirely in R, $L(M_1)$ is entirely outside R, $L(N_1)$ and $L(M_1)$ have the same inclination, and the inclination of S on both M_1P and PN_1 is monotone. See figure 2.

The simple closed curve consisting of the ray $L(M_1)$, the arc M_1PN_1, and the ray $L(N_1)$ divides the plane into two simply connected regions; let R_2 be that one containing the arc $(-\infty, M_1)$ of S. By Lemma 1 we deduce easily that R_2 contains a simple flow. Moreover, by the same lemma, $V(Q, R_2)$

exists for interior points Q of M_1PN_1. We shall prove that there is a subarc MPN of M_1PN_1, such that

$$V(N, R_2) > V(M, R_2). \tag{9}$$

Assuming for the moment that this is true, then the hypotheses of Theorem 2 are satisfied by R and R_2 (compare figures 1 and 2). It follows that $V(N, R) > V(M, R)$, proving the theorem.

Before proving (9) we shall introduce a useful notation: any region which is congruent to a given region R will be denoted by R'. If a point in R', when considered as a geometrical entity of the configuration R', corresponds under the congruence to a point Q of R, it will be denoted by Q', etc.

Now we can place (or more precisely, construct) a replica R_2' of R_2 such that R_2' is contained in R_2 and some interior point M' of $M_1'P'$ coincides with some interior point N of PN_1. Then by Theorem 1, $V(N, R_2) > V(M', R_2')$ $= V(M, R_2)$, proving (9).

Part II.

3. The problem of the wake. Consider a fixed obstacle (arc) AB immersed in a uniform flow (§ 2) having a velocity U at infinity. It is assumed that the flow separates from AB at its endpoints, leaving a *wake* or *cavity* behind AB. The streamlines Σ and Σ^* which bound the cavity are not given geometrically, but by the condition that the flow speed on Σ and Σ^* is to be constant; this constant will be called the *cavity speed.* The *problem of the wake* is to determine such a flow past the given obstacle AB (a fuller discussion of this well known problem is readily available in nearly every treatise on hydrodynamics). We shall consider the uniqueness problem for two characteristically different wakes that an obstacle might produce.

First we consider the usual *Helmholtz wake problem.* That is, Σ and Σ^* are required to extend downstream to infinity; it is well known that this can occur only when the cavity speed is U.

The second problem is the so-called *finite cavity problem.*[6] As the name implies, it is required that the cavity be finite in extent, so that i) the free streamlines meet and end at a point E behind the obstacle, and ii) the point at infinity is an interior point of the flow.

Let the flow take place in the complex z-plane. There is obviously a stagnation point O of the flow on AB; O will be taken as origin and the positive x-axis in the direction of the uniform flow. For the purposes of this

[6] A fuller discussion of the finite cavity problem is contained in [3].

paper *it is necessary to restrict our consideration henceforth to symmetric flows past symmetric obstacles.* Because of this we need only look at the "upper half" of the complete flow and, indeed, this point of view will be adopted consistently. We shall assume that AB has a continuous tangent; the inclination of the tangent vector will be denoted by the variable Ψ.[7] A symmetric obstacle will be called a *regular obstacle* if there exists a number α, $\pi/2 \leq \alpha \leq \pi$, such that $\alpha - \pi \leq \Psi \leq \alpha$ holds along OB.

Regular obstacles will be our main interest. Clearly a line whose inclination is α intersects the arc OB of a regular obstacle in at most a single point or along a straight segment. The following lemma shows that Σ enjoys the same property.

LEMMA 2. *In a symmetric solution of either the Helmholtz problem or the finite cavity problem for a regular obstacle, the inclination θ of Σ satisfies $\alpha - \pi < \theta < \alpha$.*

A proof of Lemma 2 is given in [3]. The following result is well known.

LEMMA 3. *The free streamline Σ in a symmetric solution of the Helmholtz problem may be represented parametrically by the equations*

$$x = x(\zeta) = a[\zeta^{-2} + \omega'(0)^2 \log |\zeta| + P_1(\zeta^2)],$$

$$y = y(\zeta) = a[2\omega'(0)\zeta^{-1} + P_2(\zeta^2)],$$

where $-1 < \zeta \leq 0$. Here a and $\omega'(0)$ are certain constants, and $P_1(\zeta)$ and $P_2(\zeta)$ are analytic functions in the unit circle. Hence the free streamlines are analytic curves.

Remark. If $\omega'(0) \neq 0$, the free streamlines approach the parabola $y^2 = [4a\omega'(0)^2]x$. If $\omega'(0) < 0$, Σ approaches the upper branch, and Σ^* the lower. If $\omega'(0) > 0$, Σ approaches the lower branch, and Σ^* the upper. This overlapping of the flow region in the latter case, although physically unrealistic, is not objectionable mathematically. If $\omega'(0) = 0$, then the free streamlines asymptotically approach a curve $y^2 =$ const. x^{-2n-1}, where n is a positive integer.

In the finite cavity problem a corresponding lemma exists, and the free streamlines are analytic curves. We note finally that the "upper half" of a flow solving either of our problems is a simple flow. This fact is important in the application of Theorems 1 and 2.

---

[7] In determining Ψ, the tangent vector is considered to be directed from A to B; inclination is measured in a counterclockwise sense from the positive x-axis.

4. The Helmholtz problem. We recall that it is only necessary to consider the "upper half" of a flow.

THEOREM 4. *There can be no more than one symmetric flow solving the Helmholtz problem for a regular obstacle AB.*

Proof. Suppose for contradiction that there exist two flows solving this problem. Let R_1 and R_2 be the regions occupied by these flows, and let Σ_1 and Σ_2 be the corresponding free streamlines. It is apparent by Lemma 3 that one of the free streamlines, say Σ_1, is below the other at infinity.

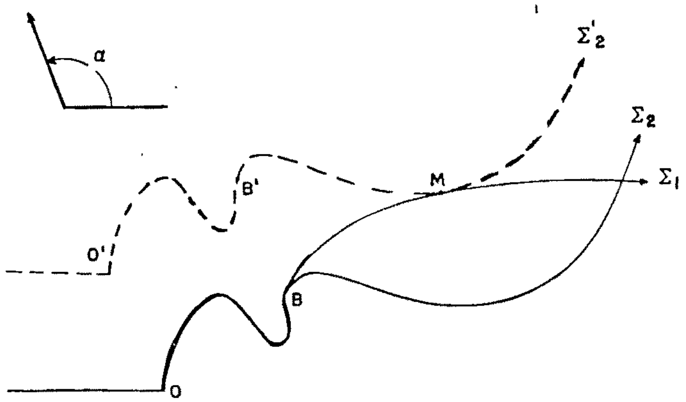


FIGURE 3

Figure 3 illustrates the situation arising from a suitable translation of a replica R_2' of R_2. (In figures, we represent a replica of a given region by a dotted curve. The prime notation associated with a replica is explained on page ——.) Here R_1 contains R_2', and Σ_1 and Σ_2' have a common point M. It will be proved in a moment that such a translation always exists (provided only that AB is a regular obstacle). Assuming this result, however, we obtain an immediate contradiction. For on one hand, by Theorem 1, $V(M, R_1) > V(M, R_2')$, while on the other hand, $V(M, R_1) = U = V(M, R_2')$, since M is on both the free streamlines Σ_1 and Σ_2'.

To complete the proof, we begin by noting a property of the curves Σ_1 and Σ_2 which follows from Lemma 2. Namely, let Λ be the line passing through B whose inclination is α. Then any line parallel to Λ,

i) cannot intersect either Σ_1 or Σ_2 if it is on the same side of Λ as OB;

ii) intersects both Σ_1 and Σ_2 in exactly one point if it is on the other side of Λ.

Now let $z = z(s)$ be a parametric representation of Σ_2. Let $d(s)$ be the distance between the point $z(s)$ of Σ_2 and the curve Σ_1, measured along a line parallel to Λ. (We take $d(s)$ positive if the distance is measured into R_2 and negative otherwise.) In virtue of ii), the function $d(s)$ is well defined, and by Lemma 3 it is continuous. $d(s)$ is zero at B and negative at infinity; therefore, it attains a maximum value d.

If the replica R_2', originally in coincidence with R_2, is translated a distance d parallel to Λ, the required situation is evidently obtained, q. e. d.

5. The schlicht cavity problem. Given an obstacle AB, the following problem will be called the *schlicht cavity problem*: find *either* a solution of the finite cavity problem for AB *or* a solution of the Helmholtz problem for AB; in the latter case it is required that, sufficiently distant from the obstacle, the free streamlines do not overlap. Evidently a uniqueness theorem for this problem will account for the finite cavity problem as well. Before stating our main result, we prove two preliminary theorems, the second justifying the name "schlicht cavity problem."

THEOREM 5. *In a symmetric flow solving the finite cavity problem the cavity speed V is less than the speed at infinity U.*

Proof. Let R be the "upper half" flow region. Obviously there is some point M on Σ with a non-positive minimal ordinate y_M. Let R_1 be half plane above the line $y = y_M$. The uniform flow $w(z) = Uz$ in R_1 is a simple flow, and R_1 contains R. Therefore from Theorem 1,

$$V = V(M, R) < V(M, R_1) = U.$$

THEOREM 6. *In a symmetric flow solving schlicht cavity problem the free streamline Σ stays entirely above the real axis.*[8]

Proof. Let R be the "upper half" flow region. We prove separately case I, when the flow solves the Helmholtz problem, and case II, when the flow solves the finite cavity problem.

Case I. Suppose the theorem is false. Then there must be some point M on Σ with a non-positive minimal ordinate y_M. Let R_1 be the half plane above the line $y = y_M$. Then by Theorem 1, $V(M, R_1) > V(M, R)$. But this is impossible, for $V(M, R) = U = V(M, R_1)$.

Case II. Let M_1 be the point of Σ with minimal ordinate, and let N_1

[8] Except, of course, at B if the flow solves the finite cavity problem.

be the point of the arc M_1E of Σ with maximal ordinate. Let R_2 be the region above the curve consisting of the arcs $L(M_1)$, M_1N_1, and $L(N_1)$ in figure 4.

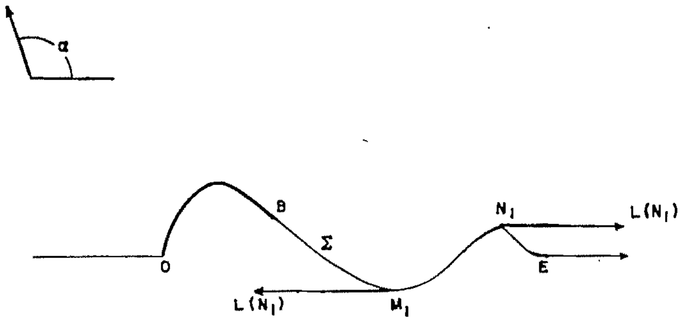


FIGURE 4

Now by Theorem 1, there exist two points M and N of the arc M_1N_1 such that $V(M,R_2) < V(N,R_2)$; (see the proof of Theorem 3). The hypotheses of Theorem 2 are satisfied by R and R_2 (compare figures 1 and 4). It follows that $V(M,R) < V(N,R)$. This is impossible since $V(M,R) = V(N,R) = V$, where V is the cavity speed.

THEOREM 7. *There can be no more than one symmetric flow solving the schlicht cavity problem for a given regular obstacle AB.*

Proof. Suppose for contradiction that there exist two flows solving this problem. Let R_1 and R_2 be the regions occupied by these flows, and let Σ_1 and Σ_2 be the corresponding free streamlines. If both flows solve the Helmholtz problem we contradict Theorem 4. There remain two cases.

Case I. The flow in R_1 solves the finite cavity problem, while the flow in R_2 solves the Helmholtz problem. Figure 5 illustrates the situation arising

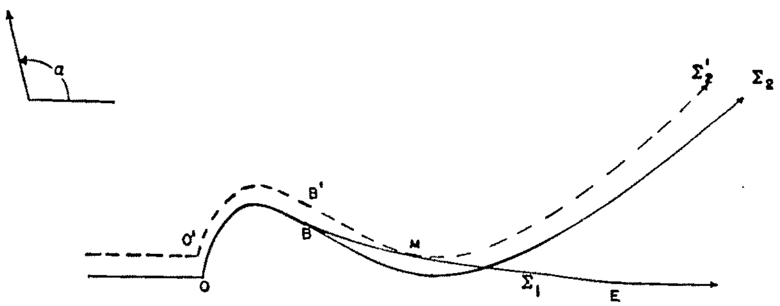


FIGURE 5

from a suitable translation of a replica R_2' of R_2. Here R_1 contains R_2', and Σ_1 and Σ_2' have a common point M. A rigorous justification of this assertion is readily obtained by the same reasoning already employed in the proof of Theorem 4, and need not be repeated.

By Theorem 1, $V(M, R_1) > V(M, R_2')$. This is impossible, however, for (Theorem 5) $V(M, R_1) = V < U$ and $V(M, R_2') = U$.

Case II. Both the flow in R_1 and the flow in R_2 solve the finite cavity problem. We shall suppose the endpoint E_1 of the streamline Σ_1 is to the right of the endpoint E_2 of the streamline Σ_2.

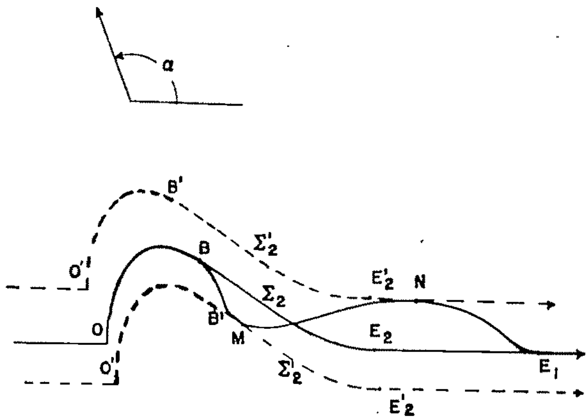


FIGURE 6

Figure 6 illustrates the situation arising from *two* suitable translations of replicas R_2' of R_2. Consider first the replica R_2' which contains R_1. Here, $V(M, R_2') > V(M, R_1)$. However, since M is on both the free streamlines Σ_2' and Σ_1, we have $V(M, R_2') = V_2$ and $V(M, R_1) = V_1$, where V_1 and V_2 are the cavity speeds. Hence

$$V_2 > V_1. \tag{10}$$

On the other hand, considering the replica R_2' inside R_1, we have

$$V(N, R_1) > V(N, R_2'). \tag{11}$$

Now N is on Σ_1, so that

$$V(N, R_1) = V_1. \tag{12}$$

However, considered as a point of R_2', N may be on either the free streamline

Σ_2' or the arc (E_2', ∞). (The latter possibility is shown in figure 6.) If N is on Σ_2' then

$$V(N, R_2') = V_2. \tag{13}$$

From (11)-(13) it follows that $V_1 > V_2$. This contradicts (10), so that N must be on (E_2', ∞). If we can show that in this case

$$V(N, R_2') > V_2, \tag{14}$$

then it will follow again that $V_1 > V_2$, contradicting (10) and completing the proof.

To prove (14), it is sufficient to show that $V(N, R_2) > V_2$, for any point N on the arc (E_2, ∞). Thus, let the region R_3 be the upper half plane. If the common arc EN of the boundaries of R_2 and R_3 is associated with the arc MN of Theorem 2, then the geometrical hypotheses of Theorem 2 are satisfied by R_2 and R_3. Moreover,

$$V(E, R_3) = V(N, R_3) = U. \tag{4'}$$

Hence $V_2 = V(E, R_2) < V(N, R_2)$. q. e. d.

6. Existence of finite cavity flows. Consider a simple closed curve T symmetric about the x-axis. Let O and O' be the two points of intersection of T with the x-axis, with O to the left of O'. Let T have the property that any symmetric arc BOB^* of T is a regular obstacle.

It can be shown [3] with the help of the Schauder-Leray fixed point theory that there exists a solution of the schlicht cavity problem for any symmetric arc BOB^* of T. If B, B^* are sufficiently near O, then any line parallel to the x-axis can intersect BOB^* in at most one point. In this case it is well known (see Leray [2], p. 157) that the (unique) solution which does exist is necessarily a Helmholtz flow. Conversely, we can guarantee that if B and B^* are sufficiently near O' the solution which does exist is necessarily a finite cavity.

Proof. Consider first the symmetric classical flow, with uniform velocity U at infinity, past the obstacle T. The real axis exterior to T is a streamline of this flow, while points O and O' are stagnation points. Since O' is a stagnation point there is some arc $NO'N^*$ of T along which the speed is everywhere less than or equal to U. We assert that if BOB^* is a symmetric arc of T and B, B^* are points of $NO'N^*$, then the solution of the schlicht cavity problem cannot be the Helmholtz flow. For, upon supposing the contrary, essentially the argument used in the proof of Theorem 7, Case I, leads to an impossibility.

7. Concluding remarks. The results of this paper carry over as well to symmetric flows past obstacles with corners. More important is the following generalization.

A curve will be called *starlike* if

i) there exists a point σ in the lower half plane $y \leq 0$, such that any straight line passing through σ intersects the curve in at most one point or along a segment, or if

ii) there exists a number α, $0 \leq \alpha \leq \pi$, such that any straight line whose inclination is α intersects the curve in at most one point or along a segment.

A symmetric obstacle with a continuous tangent (a finite number of corners may be allowed) will be called a *starlike obstacle* if the curve consisting of the negative real axis together with OB is starlike. It is apparent that the regular obstacles discussed in the body of the paper are also starlike obstacles. The proof of the following lemma is indicated in [3].

LEMMA 2′. *In a symmetric solution of the schlicht cavity problem for a starlike obstacle AB, the steramline $\psi = 0$ consisting of the negative real axis, OB, and Σ is a starlike curve.*

With the help of this lemma, Theorems 5-7 may easily be shown to hold for starlike obstacles. The proof need be modified only to the extent that our auxiliary translations of replicas of given flows regions be replaced by suitable magnification of replicas, the center of magnification being σ. Such a magnification, it is noted, leaves flow velocities unchanged at geometrically similar points.

UNIVERSITY OF INDIANA.

REFERENCES.

[1] M. Lavrentieff, "Sur certaines proprietes des fonctions univalent et leurs applications a la theorie des sillages" (in Russian), *Recueil Mathematique* (*Matematicheskii Sbornik*), vol. 46 (1938), pp. 391-458.

[2] J. Leray, "Les problemes de representation conforme d'Helmholtz; theorie des sillages et des proues," *Commentarii Mathematici Helvetici*, vol. 8 (1935-36), pp. 149 and 250.

[3] J. Serrin, "Existence theorems for some free boundary problems," *Journal of Rational Mechanics and Analysis*, vol. 1 (1952), pp. 1-48.

EXTENSIONS OF DIFFERENCE FIELDS.*

By Richard M. Cohn.[1]

1. **Introduction.** By an (abstract) difference field [4][2] we mean a field with an endomorphism. The algebraic theory of difference equations has led to the discovery of two surprising phenomena concerning difference fields. J. F. Ritt [5], § 3, showed that a difference field $\mathfrak{F}$ may have extensions $\mathfrak{G}$ and $\mathfrak{H}$ which cannot both be embedded in any one extension of $\mathfrak{F}$, and which we shall therefore call incompatible. In [1] it was shown that a difference field $\mathfrak{F}$ of characteristic 0 may have a proper extension which does not permit two distinct mappings into any extension of $\mathfrak{F}$. We call such an extension a monadic extension of $\mathfrak{F}$. A corresponding phenomenon occurs in the theory of algebraic fields, that is, fields in the ordinary sense, only if the characteristic p exceed 0. In this case an extension which is monadic in the sense just described may be produced by adjoining a p-th root not already in the field.

Our purpose in this paper is to present in organized form the still rudimentary theory of these phenomena, and to point out their decisive importance in the algebraic theory of difference equations. They seem, in fact, to mark a point byond which one can no longer use the theory of polynomial ideals or the algebraic theory of differential equations as a guide to the study of difference equations, but must expect phenomena which are *sui generis*. Our results also suggest interesting questions concerning field structure. It would, for example, be desirable to characterize those fields which possess incompatible or monadic extensions. Our principal result is a first step toward such a characterization; *fields which are algebraically closed have no finite extensions of these types.*

2. **Definitions.** By the (first) transform e_1 of an element e of a difference field $\mathfrak{F}$ we mean the element into which e is carried by the endomorphism in virtue of which $\mathfrak{F}$ is a difference field. By the transform of order k, or the k-th transform e_k of e we mean, of course, the element into which e is carried by the k-th iterate of the endomorphism (hence $e_0 = e$).

* Received April 28, 1951.

[1] This paper was written while the author held a university research grant from Rutgers University.

[2] Numbers in square brackets refer to papers listed in the Bibliography.

For the difference field $\mathfrak{G}$ to be an extension of the difference field $\mathfrak{F}$, it must be an extension of $\mathfrak{F}$ as an algebraic field, and the operation of taking transforms in $\mathfrak{G}$ must coincide for elements in $\mathfrak{F}$ with that of taking transforms in $\mathfrak{F}$. The difference field resulting from the adjunction of a set S of elements of $\mathfrak{G}$ to $\mathfrak{F}$ is by definition the least extension of $\mathfrak{F}$ containing the elements of S and their transforms of all orders. Such an extension will be called a finite (countable, etc.) extension of the difference field $\mathfrak{F}$ if S is finite (countable, etc.). Hence a finite extension of the difference field $\mathfrak{F}$ may well be an infinite extension of $\mathfrak{F}$ considered as an algebraic field. The extension just described will be denoted by $\mathfrak{F}\langle S\rangle$, while the algebraic field resulting from the adjunction of the elements of S without their transforms will be denoted by $\mathfrak{F}(S)$.

We define two extensions $\mathfrak{G}$ and $\mathfrak{H}$ of a difference field $\mathfrak{F}$ to be *incompatible* if there does not exist an extension $\mathfrak{K}$ of $\mathfrak{F}$ containing a subfield $\mathfrak{G}_1$ isomorphic to $\mathfrak{G}$, and a subfield $\mathfrak{H}_1$ isomorphic to $\mathfrak{H}$, these isomorphisms leaving each element of $\mathfrak{F}$ fixed.[3]

Let $\mathfrak{G}$ be an extension of $\mathfrak{F}$ such that not every element of $\mathfrak{G}$ has a transform of some order in $\mathfrak{F}$. We then define $\mathfrak{G}$ to be a *monadic extension* of the difference field $\mathfrak{F}$, if there does not exist an extension $\mathfrak{K}$ of $\mathfrak{F}$ into which $\mathfrak{G}$ has two distinct isomorphisms both preserving the elements of $\mathfrak{F}$.[4]

All fields occurring henceforth in this paper will be supposed to be of characteristic 0.

3. Examples and applications. *Example* 1. Let $\mathfrak{F}$ be the difference field consisting of the real numbers, with the identity as endomorphism. Let $\mathfrak{G}$ be the difference field consisting of the complex numbers, with the identity as endomorphism, and let $\mathfrak{H}$ be the difference field consisting of the complex numbers, with the transform of a number defined to be its complex conjugate. Since an extension of $\mathfrak{F}$ can contain at most one subfield isomorphic as an algebraic field to the field of complex numbers, it is evident that $\mathfrak{G}$ and $\mathfrak{H}$ are incompatible.

Let us consider the systems of difference polynomials[5] with coefficients in $\mathfrak{F}$

[3] Henceforth, when we refer to isomorphisms of extensions of a difference field $\mathfrak{F}$, it will be understood that we mean isomorphisms which leave each element of $\mathfrak{F}$ fixed.

[4] In other words no extension of $\mathfrak{F}$ can contain two distinct subfields each isomorphic to $\mathfrak{G}$ by isomorphisms which leave every element of $\mathfrak{F}$ fixed, nor does $\mathfrak{G}$ have an automorphism, other than the identity, which leaves each element of $\mathfrak{F}$ fixed.

[5] We deal here with a difference ring $\mathfrak{F}\{y\}$ consisting of all polynomials in

A) $y^2+1,\ y_1-y$, and B) $z^2+1,\ z_1+z$.

It is evident that A) has a solution, the square root of -1, which generates $\mathfrak{G}$ when adjoined to $\mathfrak{F}$, while B) has a solution, again the square root of -1, which generates $\mathfrak{H}$. Hence it is possible to solve the system A) or the system B), but not to solve these systems simultaneously, even though they have no unknown in common.

We may put the matter in a more basic, if less striking, form by saying that the manifold of the algebraically irreducible difference polynomial y^2+1 consists of two irreducible manifolds [6] $\mathfrak{M}_1$ and $\mathfrak{M}_2$. $\mathfrak{M}_1$ is the manifold of the reflexive prime difference ideal Σ_1 with characteristic set [7] $y^2+1,\ y_1-y$, and $\mathfrak{M}_2$ is the manifold of the reflexive prime difference ideal Σ_2 with characteristic set $y^2+1,\ y_1+y$. No extension of $\mathfrak{F}$ contains generic zeros [7] or, indeed, solutions of both Σ_1 and Σ_2. As suggested by this example, we define a set of reflexive prime difference ideals (or their manifolds) to be *incompatible* if there exists no extension of the coefficient field containing generic zeros of all of them. We say that an algebraically irreducible difference polynomial has an *inconsistent general solution*, if the manifolds of its general solution (see [1], p. 152) are incompatible. We shall later see that if a

---

$y_0=y,\ y_1,\ y_2\cdots$, with coefficients in $\mathfrak{F}$. The endomorphism of the ring is defined by the stipulations that elements of $\mathfrak{F}$ shall have the same transforms as in $\mathfrak{F}$ itself, and that the transform of y_i shall be y_{i+1}, $i=0,1,\cdots$. To solve a system of polynomials of $\mathfrak{F}\{y\}$, we must find a difference field $\mathfrak{G}$ which is an extension of $\mathfrak{F}$, and contains an element a such that when y is replaced by a (which means, of course, that, for each i, y_i is replaced by the i-th transform of a), the polynomials vanish. The extension of these concepts to rings in two indeterminates y, z, or rings in n indeterminates, is obvious. In the latter case we shall generally designate the indeterminates by $y_1, y_2,\cdots, y_n$. Then y_{ij} will denote the j-th transform of y_i, $i=1,\cdots,n$; $j=0,1,\cdots$.

[6] The manifold of a system of difference polynomials is the set of its solutions in all extensions of the coefficient field. It is proved in [3] that every such manifold of solutions is the union of a finite number of irreducible manifolds, that is of manifolds which, if they annul a product AB of difference polynomials, annul one of the factors. The statement in [3] does not refer to abstract difference fields, but the work of [4] makes it evident that the result applies to them also. The set of all polynomials vanishing on an irreducible manifold forms a reflexive prime difference ideal. Conversely, the solutions of a reflexive prime difference ideal constitute an irreducible manifold. This correspondence is one-to-one.

[7] Characteristic sets of difference polynomials are defined in [3], where they are called basic sets. In this paper we have brought the terminology for difference polynomials into accord with the new terminology for differential polynomials adopted by Ritt in [6]. We list here the terms used in earlier papers, each followed by the term which is to replace it: arbitrary unknown, parametric indeterminate; ascending set, chain; basic sequence, characteristic sequence; basic set, characteristic set; general point, generic zero.

difference field has a finite extension which is incompatible with some other extension, then there is an algebraically irreducible zero-order difference polynomial with coefficients in the field, whose general solution is inconsistent.

As a result of the existence of incompatible extensions, the dimension or order [8] of a reflexive prime difference ideal may be reduced when the coefficient field is extended. Thus the system A), which is of zero order in $\mathfrak{F}$, has no solution if $\mathfrak{H}$ is taken as the coefficient field. The system

$$\text{C)} \qquad y_k^2 + y^3, \qquad y_{k+1}y - y_k y_1,$$

is the characteristic set of a prime ideal Σ of order k with coefficients in $\mathfrak{F}$, as we know from Theorem IV of [1]. If α is any solution of Σ except 0, then since α satisfies C), α_k/α is a solution of A). Hence $\mathfrak{F}\langle\alpha\rangle$ is an extension of $\mathfrak{G}$. It follows that 0 is the only solution [9] of the polynomials of Σ, considered as a system of polynomials with coefficients in $\mathfrak{H}$. Hence the order of this system is reduced from k to 0 when the coefficient field is extended to $\mathfrak{H}$.

Incompatible extensions also lead to difficulties when we try to replace parametric indeterminates by special values. From the analogy with systems of algebraic or algebraic differential equations, we should expect that if Π is a reflexive prime difference ideal in the indeterminates $u_1, u_2, \cdots, u_q$; $y_1, y_2, \cdots, y_p$, the u_i being parametric indeterminates,[10] it would be possible to find a solution of Π in which the u_i have any assigned values not annulling a certain difference polynomial U in the u_i. But this is not true for the reflexive prime ideal Π in indeterminates [11] u, y, with characteristic set

$$\text{D)} \qquad u^2y^2 + 1, \qquad u_1y_1 + uy.$$

Here u is a set of parametric indeterminates for Π, and the coefficient field is $\mathfrak{F}$ above. Suppose a U of order m exists. Let α be a generic zero of a system C) with $k > m$. Then Theorem IV of [1] assures us that α does

[8] "Dimension" corresponds to the heuristic concept of the number of indeterminates whose values can be chosen arbitrarily (parametric indeterminates). "Order" corresponds to the heuristic concept of the number of arbitrary periodic functions in the solution. See [1], p. 141 and p. 162.

[9] That 0 is a solution of Σ can be shown by an easy application of the technique used in our paper "Singular manifolds of difference polynomials," *Annals of Mathematics*, vol. 53, pp. 445-463. Let ϕ be a generic zero of Σ. Adjoin to $\mathfrak{F}$ an element c which is its own transform, and which satisfies no algebraic equation over $\mathfrak{F}$. Then ϕc annuls the polynomials of C) but not their initials. Hence ϕc annuls the polynomials of Σ. It follows readily from this that 0 is in the manifold of Σ.

[10] Here single subscripts do not indicate transforms.

[11] Single subscripts do indicate transforms here.

not annul U. Now $\mathfrak{F}\langle\alpha\rangle$ is an extension of $\mathfrak{G}$, while every solution of II lies in an extension of $\mathfrak{H}$. Hence there cannot be a solution of II with $u = \alpha$. Thus the assumption that U exists leads to a contradiction.

The existence of incompatible extensions also has implications for the analytic theory of difference equations. We consider difference polynomials in a field $\mathfrak{Z}$ of functions of the complex variable z meromorphic on a line $\mathfrak{L}$ which contains $z+1$ if it contains z. The transform of a function $f(z)$ is defined to be $f(z+1)$. If S_1 and S_2 are systems of such difference polynomials with the property that no extension of $\mathfrak{Z}$ as an abstract difference field contains both a solution of S_1 and a solution of S_2, then obviously at most one of these systems has solutions meromorphic on $\mathfrak{L}$. Of the systems A) and B) the latter has, of course, no meromorphic solutions whatsoever.

Example 2. We consider the field of rational functions of x with complex coefficients. This becomes a difference field $\mathfrak{F}$, if we define the transform of any rational function $f(x)$ to be the function $f(4x)$. Consider the difference polynomial $A = y^2 - x$ which is algebraically irreducible in $\mathfrak{F}$. Its transform A_1 is $y_1{}^2 - 4x$. Then $A_1 - 4A = y_1{}^2 - 4y^2 = (y_1 - 2y)(y_1 + 2y)$. Hence every solution of A annuls either

E) $\quad y^2 - x, y_1 - 2y$ or $\qquad$ F) $\quad y^2 - x, y_1 + 2y$.

We know from Theorem IV of [1] that both E) and F) have solutions in suitable extensions of $\mathfrak{F}$. We shall show that no one extension contains solutions of both, so that A has an inconsistent general solution.

Let α be a solution of E). The only other solution of A lying in an extension of $\mathfrak{F}\langle\alpha\rangle$ is $\beta = -\alpha$. From E) we have $\alpha_1 = 2\alpha$. Now $\beta_1 = -\alpha_1 = -2\alpha = 2\beta$. Hence β annuls E), and no extension of $\mathfrak{F}\langle\alpha\rangle$ contains a solution of F).

Example 3. We now define the transform of a rational function $f(x)$ with complex coefficients to be $f(x^3)$, and let $\mathfrak{F}$ be the difference field consisting of all such functions of x with transforms defined in this way. We find that every solution of the algebraically irreducible difference polynomial $y^2 - x$ annuls either

G) $\quad y^2 - x, y_1 - yx$ or $\qquad$ H) $\quad y^2 - x, y_1 + yx$.

As before we see that the manifolds of G) and of H) are incompatible.

Example 4. If we define the transform of a rational function $f(x)$ with complex coefficients to be $f(x^2)$, we get a difference field $\mathfrak{F}$ which has monadic

extensions. For example, every solution of the algebraically irreducible difference polynomial $y^3 - x$ annuls one of the systems

I) $y^3 - x, y_1 - y^2$, J) $y^3 - x, y_1 - \omega y^2$,

K) $y^3 - x, y_1 - \omega^2 y^2$,

where ω is a complex cube root of unity. All these systems have solutions in suitable extensions of $\mathfrak{F}$. Let α be a solution, say of I). Then $\omega\alpha$, $\omega^2\alpha$, are the only other solutions of $y^3 - x$ in any extension of $\mathfrak{F}\langle\alpha\rangle$. It is easy to see that they annul the systems K) and J) respectively. Hence no extension of $\mathfrak{F}\langle\alpha\rangle$ contains a subfield other than $\mathfrak{F}\langle\alpha\rangle$ itself which is isomorphic to $\mathfrak{F}\langle\alpha\rangle$ by an isomorphism leaving each element of $\mathfrak{F}$ fixed; nor does $\mathfrak{F}\langle\alpha\rangle$ have any automorphism except the identity which leaves each element of $\mathfrak{F}$ fixed. No transform of α is in $\mathfrak{F}$. Hence $\mathfrak{F}\langle\alpha\rangle$ is a monadic extension of $\mathfrak{F}$. J) and K) likewise furnish monadic extensions of $\mathfrak{F}$.

It is interesting to note that if we define the transform of a rational function $f(x)$ with complex coefficients to be $\bar{f}(x^2)$, where $\bar{f}$ is the function obtained by replacing each coefficient in f by its complex conjugate, we obtain a difference field $\mathfrak{G}$ in which the systems I), J), K) define not monadic extensions, but rather extensions any two of which are incompatible.

We define a reflexive prime difference ideal Σ, not containing 1, or its manifold $\mathfrak{M}$ to be monadic, if no transform of a generic zero of Σ lies in the coefficient field, and if there exists no extension of the coefficient field containing two generic zeros of Σ. We shall see that if a difference field $\mathfrak{F}$ has finite monadic extensions (and consequently monadic prime ideals), it has a monadic extension which is obtained by adjoining a solution of a zero order difference polynomial in one indeterminate. In fact, the situation with finite monadic extensions is even stronger than with finite incompatible extensions—every finite monadic extension is of zero effective order.[12]

The proof that finite algebraic extensions of a field of characteristic zero are simple, and the proof of the corresponding theorem for differential fields,[13] are based on the fact that such fields do not have monadic extensions. For fields of characteristic p the proof fails and the theorem does not hold. In the case of difference fields of characteristic zero, however, the difficulty

[12] The concept of a quasi-linear system, defined in [1], p. 164, includes monadic prime ideals and the uninteresting prime ideals whose generic zeros or their transforms lie in the coefficient field. The result we have just referred to is essentially Theorem XI of [1], which we prove again in slightly altered form in the corollary to Theorem II below.

[13] See [6], pp. 41 and 86.

caused by the existence of monadic extensions can be circumvented, and a completely analogous theorem can be proved.[14] The corollary to Theorem II of this paper (or Theorem XI of [1]) is required for the proof.

4. Preliminary lemmas. As stated above we shall see that all difference fields which have finite incompatible extensions, or a finite monadic extension, have incompatible or monadic extensions of the types shown in the examples; i. e., extensions formed by adjoining solutions of a zero order difference polynomial in one indeterminate. For both incompatible and monadic extensions the proof depends essentially on Lemma III of this section, which in turn depends on Lemma II. It is convenient before considering these lemmas to make some conventions regarding notations.

Let a set of indeterminates $y_1, \cdots, y_n$ be adjoined to a difference field $\mathfrak{F}$. The least difference ring containing $\mathfrak{F}$ and these indeterminates will be denoted by $\mathfrak{F}\{y_1, \cdots, y_n\}$. We denote by y_{ij} the j-th transform of y_i.

Let S be a set of difference polynomials of $\mathfrak{F}\{y_1, \cdots, y_n\}$. The least difference ideal containing the polynomials of S will be denoted by $[S]$, while the perfect difference ideal which they generate is indicated by $\{S\}$.

Now let $\mathfrak{F}$ be any field, not necessarily a difference field. Let indeterminates $y_1, \cdots, y_n$ be adjoined to $\mathfrak{F}$ to form a ring, which does not, of course, contain transforms of the y_i. We denote this ring by $\mathfrak{F}[y_1, \cdots, y_n]$. Let S represent a set of polynomials of this ring. The least ideal containing the polynomials of S will be denoted by $(S)_0$. For definiteness, ideals in $\mathfrak{F}[y_1, \cdots, y_n]$ will be called polynomial ideals. The set of all polynomials of which a power is in $(S)_0$ is the perfect polynomial ideal $\{S\}_0$. It may happen that $\mathfrak{F}$ is a difference field, and the indeterminates of this paragraph are a finite subset of transforms of the indeterminates of a difference ring. Then $(S)_0$ involves only these transforms.

We shall sometimes want to indicate that one of the ideals just described is to be formed from S in an extension $\mathfrak{G}$ of $\mathfrak{F}$. In such a case the symbol $\mathfrak{G}$ is placed in the brackets after S, and separated from it by a semicolon. We write, for example, $\{S; \mathfrak{G}\}$ for the perfect difference ideal generated by polynomials of S in the field $\mathfrak{G}$.

From this point on we depend heavily on the results of [1], [3], [4], and [6]. We remind the reader of the changes in terminology listed in footnote 7.

[14] The proof is given in our paper "On extensions of difference fields and the resolvents of prime difference ideals," to be published in the *Proceedings of the American Mathematical Society.*

Our first lemma is a restatement of a well-known theorem concerning algebraic, not difference polynomials.

LEMMA I. *Let* Π *be a prime polynomial ideal with coefficients in a field* $\mathfrak{F}$. *Let* $\mathfrak{G}$ *be an extension of* $\mathfrak{F}$. *Then* $\{\Pi;\mathfrak{G}\}_0$ *is the intersection of prime ideals* $\Pi_1,\Pi_2,\cdots,\Pi_r$, *which are such that any set of parametric indeterminates of* Π *is a set of parametric indeterminates of each* Π_k.

Proof. Let Π involve the indeterminates $u_1,u_2,\cdots,u_q;y_1,y_2,\cdots,y_p$, the u_i forming a set of parametric indeterminates. We form a resolvent [15] for Π using $w=\sum_{i=1}^{p}\mu_iy_i$. The μ_i may be chosen as integers. Let Ω be the prime polynomial ideal $\{\Pi, w-\sum_{i=1}^{p}\mu_iy_i\}_0$, and let

$$\text{L)} \qquad A;A_1,\cdots,A_p$$

be a characteristic set for Ω, with A introducing w, and A_i introducing y_i, $i=1,\cdots,p$. Each A_i is then of the first degree in y_i.

Let the irreducible factors of A in the field $\mathfrak{G}$ be $B_1,\cdots,B_r$. Each B_k effectively involves w, for otherwise A would not be irreducible in $\mathfrak{F}$.

We adjoin to $\mathfrak{G}$ a set of elements α_i, $i=1,\cdots,q$, which annul no non-zero polynomial with coefficients in $\mathfrak{G}$.

B_k has a solution $u_i=\alpha_i$, $i=1,\cdots,q;w=\beta_k$. This solution can annul no initial of any A_j; for these initials are not multiples of B, and hence some linear combination of an initial with B is a non-zero polynomial in the u_i alone. In each A_j we replace the u_i by the α_i, and w by β_k, forming a polynomial $A_j^{(k)}$ of first degree in y_j. By solving $A_j^{(k)}$ we find the value $\gamma_j^{(k)}$ for y_j.

Let Ω_k be the prime polynomial ideal consisting of those polynomials of $\mathfrak{G}[u_1,\cdots,u_q;w;y_1,\cdots,y_p]$ which vanish when we put $u_i=\alpha_i$, $i=1,\cdots,q$; $w=\beta_k$; $y_j=\gamma_j^{(k)}$, $j=1,\cdots,p$. Evidently Ω_k contains no polynomial in the u_i alone. We claim that $\{\Omega;\mathfrak{G}\}_0$ is the intersection of the Ω_k.

We see that the solution of Ω_k given in the preceding paragraph annuls the polynomials of the characteristic set L), but not their initials. Hence $\Omega\subset\Omega_k$, and therefore $\{\Omega;\mathfrak{G}\}_0\subset\Omega_k$.

Let P be a polynomial which is in every Ω_k. We form the remainder R of P with respect to the chain $A_1,\cdots,A_p$. R is a polynomial in the u_i and w alone, which vanishes for the solution $u_i=\alpha_i$, $i=1,\cdots,q;w=\beta_k$

[15] See [6], p. 86.

of Ω_k. Hence R is a multiple of B_k. Since this is true for each k from 1 to r, R is a multiple of A.

It follows that P has zero remainder with respect to L). We may write P as a finite sum $\sum \delta_i E_i$, where the δ_i are elements of $\mathfrak{G}$ which are linearly independent with respect to $\mathfrak{F}$, and the E_i are polynomials with coefficients in $\mathfrak{F}$. It is easy to see that each E_i has zero remainder with respect to L). Hence the E_i are in Ω, and therefore P is in $\{\Omega; \mathfrak{G}\}_0$. This proves our statement that $\{\Omega; \mathfrak{G}\}_0$ is the intersection of the Ω_k.

Now $\{\Pi; \mathfrak{G}\}_0$ consists of those polynomials of $\{\Omega; \mathfrak{G}\}_0$ which are free of w. For we may write any polynomial with coefficients in $\mathfrak{G}$ in the form $\sum \delta_i E_i$, where the δ_i are elements of $\mathfrak{G}$ which are linearly independent with respect to $\mathfrak{F}$, and the E_i are polynomials with coefficients in $\mathfrak{F}$. The set of all polynomials which have a representation of this form in which every E_i is in Π forms an ideal, and it can easily be seen that this ideal is $(\Pi; \mathfrak{G})_0$. Those polynomials having representations of this form in which every E_i is in Ω form the ideal $(\Omega; \mathfrak{G})_0$. Now Π consists of those polynomials of Ω which are free of w. A polynomial of $(\Omega; \mathfrak{G})_0$ is free of w if and only if each E_i is free of w. Hence $(\Pi; \mathfrak{G})_0$ consists of those polynomials of $(\Omega; \mathfrak{G})_0$ which are free of w. Finally $\{\Pi; \mathfrak{G}\}_0$ consists of polynomials of which a power is in $(\Pi; \mathfrak{G})_0$; and $\{\Omega; \mathfrak{G}\}_0$ consists of polynomials of which a power is in $(\Omega; \mathfrak{G})_0$. Our statement follows immediately from this.

Let Π_k be the prime polynomial ideal consisting of those polynomials of Ω_k which are free of w. Then it follows from the preceding statements that $\{\Pi; \mathfrak{G}\}_0$ is the intersection of the Π_k. The Π_k are prime ideals, and the u_i form a set of parametric indeterminates[16] for each Π_k. This proves the lemma.

LEMMA II. *With the notation of Lemma* I, *if every element of* $\mathfrak{G}$ *not contained in* $\mathfrak{F}$ *is transcendental over* $\mathfrak{F}$, *then* $\{\Pi; \mathfrak{G}\}_0$ *is itself a prime polynomial ideal, and a characteristic set of* Π *is a characteristic set of* $\{\Pi; \mathfrak{G}\}_0$. In particular, if $\mathfrak{F}$ is algebraically closed and $\mathfrak{G}$ is *any* extension of $\mathfrak{F}$, then $\{\Pi; \mathfrak{G}\}_0$ is prime, and a characteristic set of Π is a characteristic set of $\{\Pi; \mathfrak{G}\}_0$.

Proof. To show that $\{\Pi; \mathfrak{G}\}_0$ is prime, it is sufficient to show that A is irreducible in $\mathfrak{G}$; for then $k = 1$ in the proof of Lemma I. Now a poly-

[16] This is so because Ω_k contains no non-zero polynomial in the u_i alone. Hence Π_k contains no non-zero polynomial in the u_i alone. On the other hand Π, and hence each Π_k, contains a non-zero polynomial in $u_1, \cdots, u_q$ and y_j for each j from 1 to p.

nomial S in one indeterminate has the same factors in $\mathfrak{G}$ and in $\mathfrak{F}$, if we consider two factors to be identical when their quotient is an element of the coefficient field. For first of all, it is possible to restrict S and its factors to polynomials with initial coefficient 1. Secondly, S has a factorization into linear factors with initial coefficient 1 in some algebraic extension of $\mathfrak{F}$. All factors of S in any field which meet our restriction are products of these linear factors, and therefore have coefficients algebraic over $\mathfrak{F}$. Hence if such factors have coefficients in $\mathfrak{G}$, then they have coefficients in $\mathfrak{F}$.

By a device due to Kronecker[17] the factors of A can be formed from those of a related polynomial S in one indeterminate. It is easy to see from Kronecker's method of obtaining the factors of A from those of S that, since S has the same factors in $\mathfrak{G}$ as in $\mathfrak{F}$, the same is true of A. Hence A is irreducible in $\mathfrak{G}$, and $\{\Pi;\mathfrak{G}\}_0$ is a prime polynomial ideal.

Let $B_1, \cdots, B_p$ be a characteristic set of Π, and $C_1, \cdots, C_p$ a characteristic set of $\{\Pi;\mathfrak{G}\}_0$. Suppose some C_i is reduced with respect to $B_1, \cdots, B_p$. We let k be the least integer such that C_k is reduced with respect to this chain. We may then take $C_i = B_i$, $1 \leqq i < k$.

We adjoin to $\mathfrak{G}$ elements α_i, $i = 1, \cdots, q$, annulling no polynomial with coefficients in $\mathfrak{G}$. Replacing the u_i by α_i, we solve B_1 for y_1. Substituting the α_i for the u_i in B_2, and the solution just obtained for y_1, B_2 becomes a polynomial in y_2 alone. Solving this and substituting in a similar way into B_3 we obtain a polynomial in y_3 alone. Continuing in this way, we find all the y_i. At no point are we unable to solve; for each ideal $(B_1, \cdots, B_j, I_{j+1};\mathfrak{F})_0$, where I_{j+1} is the initial of B_{j+1}, contains a non-zero polynomial in the u_i alone,[18] so that I_{j+1} cannot be annulled by the solution obtained for $B_1, \cdots, B_j$. In the same way we see that a solution obtained in this manner annuls no separant of any B_j. Hence when we come to solve B_k, we can find as many distinct solutions as its degree in y_k. In particular, then, we may choose a solution of B_k which does not annul C_k. (Since the chain $B_1, \cdots, B_{k-1}$ is a characteristic set of a prime ideal of polynomials with coefficients in $\mathfrak{G}$, the initial of C_k cannot be annulled by the solution obtained for the u_i and y_j, $j < k$.)

The solution we obtain in this way annuls Π. Hence it annuls $\{\Pi;\mathfrak{G}\}_0$. This contradicts the fact that it does not annul C_k. Hence no C_i is reduced with respect to $B_1, \cdots, B_p$. This proves Lemma II.

The discussion of Example 1 shows that there is no possibility of obtaining

[17] See Van der Waerden, *Moderne Algebra*, 1st ed., vol. 1, p. 129.

[18] See [6], Chapter IV, § 9.

a complete analogue of Lemma I for difference polynomials. We can, however, obtain a result analogous to Lemma II.

LEMMA III. *Let $\mathfrak{F}$ be a difference field and Π a reflexive prime difference ideal of polynomials with coefficients in $\mathfrak{F}$. Let $\mathfrak{G}$ be an extension of $\mathfrak{F}$, which is such that every element of $\mathfrak{G}$ not contained in $\mathfrak{F}$ is transcendental over $\mathfrak{F}$, considered as an algebraic field. Then $\{\Pi; \mathfrak{G}\}$ is a reflexive prime difference ideal. Any set of parametric indeterminates of Π is a set of parametric indeterminates of $\{\Pi; \mathfrak{G}\}$, and the effective order of Π in the remaining indeterminates equals the effective order of $\{\Pi; \mathfrak{G}\}$ in those indeterminates.*

In particular, if $\mathfrak{F}$ is algebraically closed, and $\mathfrak{G}$ is *any* extension of $\mathfrak{F}$, the statements of the lemma apply.

Proof. Let $u_1, u_2, \cdots, u_q; y_1, y_2, \cdots, y_p$ be the indeterminates of Π, the u_i forming a set of parametric indeterminates.

The set of polynomials of order not exceeding k in each u_i and y_j will be denoted by Π_k, $k = 0, 1, \cdots$. Π_k is not empty for k sufficiently large, and it is a prime polynomial ideal in the indeterminates u_{im}, $i = 1, \cdots, q; m = 0, \cdots, k$, and y_{jn}, $j = 1, \cdots, p; n = 0, \cdots, k$.

By Lemma II, the polynomial ideal $\Lambda_k = \{\Pi_k; \mathfrak{G}\}_0$ is evidently a prime ideal in the same set of indeterminates as Π_k. The union Λ of the Λ_k is a prime ideal in the infinite set of indeterminates u_{im}, y_{jn}, where now the second subscripts have the range $0, 1, \cdots$. We shall show that Λ is actually a difference ideal in the u_i and y_j.

Let C be in Λ. We write C as $\sum \gamma_i C_i$, where the γ_i are a finite number of elements of $\mathfrak{G}$ linearly independent with respect to $\mathfrak{F}$, and the C_i are difference polynomials with coefficients in $\mathfrak{F}$.

Now C has zero remainder with respect to the characteristic set of some Λ_k. By Lemma II we may use a characteristic set of Π_k as a characteristic set of Λ_k. Then each C_i has zero remainder with respect to this characteristic set. Hence each C_i is in Π_k, it first transform is in Π_{k+1}, and the transform of C is in Λ_{k+1}, and therefore in Λ.

We shall show that the reflexive prime difference ideal $\{\Lambda\}$ is $\{\Pi; \mathfrak{G}\}$. On the one hand, since every Π_k is in Π, every Λ_k is in $\{\Pi; \mathfrak{G}\}$. Hence $\Lambda \subset \{\Pi; \mathfrak{G}\}$, and therefore $\{\Lambda\} \subset \{\Pi; \mathfrak{G}\}$. On the other hand, since $\Pi \subset \Lambda$, $\{\Pi; \mathfrak{G}\} \subset \{\Lambda; \mathfrak{G}\} = \{\Lambda\}$.

No Λ_k contains a non-zero polynomial in the u_{im} alone. Hence the u_i constitute a set of parametric indeterminates for $\{\Lambda\}$. It remains to show that $\{\Lambda\}$ and Π are of equal effective order in the y_i.

Let A_j be that leader of a characteristic set of Π which introduces y_j. By the result of [1], p. 149, we see that Π is held by no non-zero polynomial in y_1 and the u_i alone which is of lower effective order than A_1 in y_1. By Theorem IX of [1] and the result of [1], p. 149, we see that Π is held by no non-zero polynomial in y_1, y_2 and the u_i, of lower effective order than A_1 in y_1, and of lower effective order than A_2 in y_2. Continuing, we have the following result: Let T designate the set of y_{jn} such that, for each j, n takes those values, and only those, which are less than the effective order of A_j in y_j, and let T_g designate the set consisting of the g-th transforms of the y_{jn} in T. Then Π contains no non-zero difference polynomial in the u_{im} and the y_{jn} of any T_g. By [1], Theorem X, the number of y_{jn} in T is the effective order of Π in the y_j. To prove the lemma it suffices to show that $\{\Lambda\}$ contains no non-zero difference polynomial in the y_{jn} of any T_g and u_i.

No Π_k contains a polynomial in u_{im} and the y_{jn} of a T_g. Hence for all sufficiently large values of k, T_g and the u_{im} present in Π_k form part of a set of parametric indeterminates for Π_k. By Lemma II, they form part of a set of parametric indeterminates for Λ_k. Hence Λ contains no non-zero polynomial in any set T_g and u_i. Then $\{\Lambda\}$ contains no non-zero polynomial in any set T_g and u_i. This proves Lemma III.

We shall need one more result of a different sort, analogous to Gourin's Theorem for differential polynomials.[19]

LEMMA IV. *Let Π and Π' be reflexive prime difference ideals such that Π is properly contained in Π'. Then either Π' is of lower dimension than Π, or it is of the same dimension as Π, every parametric set for Π' is a parametric set for Π, and Π' is of lower effective order than Π relative to any parametric set.*

Proof. Since Π contains no non-zero polynomial in a set of parametric indeterminates of Π', only the last statement of the lemma requires any consideration. Let us assume then that $u_1, \cdots, u_q$ is a parametric set for both Π and Π', the remaining indeterminates being $y_1, \cdots, y_p$.

For non-negative integers a, b, c, let Π_{abc} represent the prime polynomial ideal consisting of all difference polynomials in Π which involve no transform of any y_j of order less than a or more than $a+b$, and no transform of any u_i of order more than c. For a, b, and c sufficiently large, the dimension of Π_{abc} is the sum of the effective order of Π in the y_j and the number $(c+1)q$ of u_{im} with $m \leqq c$.

[19] E. Gourin, "On irreducible systems of algebraic differential equations," *Bulletin of the American Mathematical Society*, vol. 39 (1933), pp. 593-595.

Let polynomial ideals Π'_{abc} be defined for Π' as the Π_{abc} were defined for Π. For all sufficiently large a, b, and c, the dimension of Π'_{abc} is the sum of the effective order of Π' and $(c+1)q$. Now Π'_{abc} contains Π_{abc}. Π' contains a polynomial P which is not in Π. For all sufficiently large a, b, and c, Π'_{abc} contains a transform of P, and hence contains Π_{abc} as a proper subset. Then Π'_{abc} is of lower dimension than Π_{abc} when a, b, and c are large. The lemma follows from the relation above between effective order and dimension.

Corollary. *Under the hypotheses of Lemma* III, $\{\Pi;\mathfrak{G}\}$ *contains no polynomial with coefficients in* $\mathfrak{F}$ *except the polynomials of* Π.

Proof. The polynomials of $\{\Pi;\mathfrak{G}\}$ with coefficients in $\mathfrak{F}$ constitute a reflexive prime difference ideal Λ. Since $\Pi \subset \Lambda$, evidently $\{\Pi;\mathfrak{G}\} = \{\Lambda;\mathfrak{G}\}$. By Lemma III, a set of parametric indeterminates for Π is a set of parametric indeterminates for $\{\Pi;\mathfrak{G}\} = \{\Lambda;\mathfrak{G}\}$, and hence is part of a set for Λ; by another application of Lemma III, it is a set for Λ. The effective order of Π relative to this parametric set equals that of Λ by the same lemma. Then, by Lema IV, Λ cannot contain Π as a proper subset. Hence Π and Λ are identical.

5. Incompatible extensions.

Theorem I. *Let the difference field* $\mathfrak{F}$ *have the incompatible extensions* $\mathfrak{G}$ *and* $\mathfrak{H}$, *of which at least one is a finite extension. Then there is an algebraically irreducible difference polynomial of zero order with coefficients in* $\mathfrak{F}$ *whose general solution is inconsistent.*

Hence, if $\mathfrak{F}$ is algebraically closed it has no incompatible extensions one of which is finite.

Proof. Let $\mathfrak{G}$, say, be a finite extension of $\mathfrak{F}$, and let it arise by the adjunction of elements $\alpha_1, \alpha_2, \cdots, \alpha_r$ to $\mathfrak{F}$. Let Σ be the set of all difference polynomials in unknowns $y_1, y_2, \cdots, y_r$, with coefficients in $\mathfrak{F}$, which are annulled when each y_i is replaced by the corresponding α_i. Σ is not the ideal $[0]$, for otherwise we could find an extension of $\mathfrak{H}$ containing a subfield isomorphic to $\mathfrak{G}$ merely by adjoining to $\mathfrak{H}$ r elements annulling no algebraic difference polynomial with coefficients in $\mathfrak{G}$. Evidently Σ is a reflexive prime difference ideal.

In terms of the ideal Σ we have the following condition for an extension

$\mathfrak{K}$ of $\mathfrak{F}$ to be incompatible with $\mathfrak{G}$: $\mathfrak{K}$ is incompatible with $\mathfrak{G}$ if and only if $\{\Sigma; \mathfrak{K}\}$ contains a polynomial with coefficients in $\mathfrak{F}$ which is not in Σ.

For suppose, first, that $\mathfrak{K}$ and $\mathfrak{G}$ are not incompatible. Then an extension $\mathfrak{K}'$ of $\mathfrak{K}$ contains a subfield isomorphic to $\mathfrak{G}$. Let $\bar{\alpha}_i$ be the element of $\mathfrak{K}'$ corresponding to α_i. The $\bar{\alpha}_i$ must annul the polynomials of Σ, but no other polynomial with coefficients in $\mathfrak{F}$. Since every solution of Σ in an extension of $\mathfrak{K}$ is a solution of $\{\Sigma; \mathfrak{K}\}$, it follows that $\{\Sigma; \mathfrak{K}\}$ contains no polynomials with coefficients in $\mathfrak{F}$ except the polynomials of Σ.

Next suppose that $\{\Sigma; \mathfrak{K}\}$ contains no polynomial with coefficients in $\mathfrak{F}$ except the polynomials of Σ. Let $\Pi_1, \Pi_2, \cdots, \Pi_r$ be the reflexive prime difference ideals whose intersection is $\{\Sigma; \mathfrak{K}\}$. At least one Π_i contains no polynomial with coefficients in $\mathfrak{F}$, except the polynomials of Σ. For otherwise there would exist for each i, $i = 1, \cdots, r$, a polynomial B_i of Π_i, with coefficients in $\mathfrak{F}$, which is not in Σ. The product of the B_i would be a polynomial with coefficients in $\mathfrak{F}$ which is in $\{\Sigma; \mathfrak{K}\}$ but not in Σ. This would contradict our assumption. Let Π_1, say, contain no polynomials with coefficients in $\mathfrak{F}$ except the polynomials of Σ. We adjoin a generic zero of Π_1 to $\mathfrak{K}$ forming $\mathfrak{K}'$. The subfield of $\mathfrak{K}'$ formed by adjoining the generic zero of Π_1 to $\mathfrak{F}$ is isomorphic to $\mathfrak{G}$. Hence $\mathfrak{K}$ and $\mathfrak{G}$ are not incompatible. This proves our statement of the condition for incompatibility.

Returning to the consideration of the extension $\mathfrak{H}$, we let $\mathfrak{H}^*$ consist of those elements of $\mathfrak{H}$ which are algebraic over $\mathfrak{F}$. We claim that $\mathfrak{H}^*$ and $\mathfrak{G}$ are incompatible.

Suppose that $\mathfrak{H}^*$ and $\mathfrak{G}$ are not incompatible. Then, by the condition just given, $\{\Sigma; \mathfrak{H}^*\}$ contains no polynomials with coefficients in $\mathfrak{F}$ except the polynomials of Σ. It follows as before that, among the reflexive prime difference ideals $\Pi_1, \cdots, \Pi_r$ whose intersection is $\{\Sigma; \mathfrak{H}^*\}$, at least one, say Π_1, contains no polynomial with coefficients in $\mathfrak{F}$ except the polynomials of Σ.

Let Λ denote $\{\Pi_1; \mathfrak{H}\}$. The corollary to Lemma IV shows that Λ contains no polynomial with coefficients in $\mathfrak{H}^*$, except the polynomials of Π_1. Hence Λ contains no polynomial with coefficients in $\mathfrak{F}$, except the polynomials of Σ. Since $\{\Sigma; \mathfrak{H}\} \subset \Lambda$, our criterion for incompatibility shows that $\mathfrak{G}$ and $\mathfrak{H}$ are not incompatible. This contradiction shows that $\mathfrak{G}$ and $\mathfrak{H}^*$ are incompatible.

$\mathfrak{H}^*$ may be an infinite extension of $\mathfrak{F}$, but $\mathfrak{G}$ must be incompatible with a finite zero-order extension[20] of $\mathfrak{F}$. For $\{\Sigma; \mathfrak{H}^*\}$ contains, by the above

[20] A finite zero-order extension of $\mathfrak{F}$ is an extension of the form $\mathfrak{F}\langle\gamma_1, \cdots, \gamma_n\rangle$,

criterion, a polynomial B with coefficients in $\mathfrak{F}$ which is not in Σ. As shown in [4], p. 449, B can be obtained from a finite number of polynomials of Σ by a finite number of shufflings.[21] These operations involve only a finite number of elements of $\mathfrak{H}^*$. Let $\mathfrak{H}'$ be the difference field formed by adjoining these elements to $\mathfrak{F}$. Then $\{\Sigma; \mathfrak{H}'\}$ contains B. Hence $\mathfrak{G}$ and $\mathfrak{H}'$ are incompatible.

$\mathfrak{H}'$ is a finite zero order extension of $\mathfrak{F}$. But this implies that it is a simple extension. For let $c_1, \cdots, c_s$ be the generators of $\mathfrak{H}'$. There is an element c, a linear combination with integer coefficients of the c_i, such that the algebraic field $\mathfrak{F}(c)$ contains the c_i. Then the difference field $\mathfrak{F}\langle c\rangle$ contains the c_i, and hence is $\mathfrak{H}'$.

Since $\mathfrak{H}'$ is a finite extension incompatible with $\mathfrak{G}$, a repetition of the preceding argument shows that $\mathfrak{G}$ contains an element d algebraic over $\mathfrak{F}$ such that $\mathfrak{G}' = \mathfrak{F}\langle d\rangle$ is incompatible with $\mathfrak{H}'$.

We now know that the existence of the incompatible extensions $\mathfrak{G}$ and $\mathfrak{H}$ implies the existence of simple extensions algebraic over $\mathfrak{F}$ which are incompatible. It only remains to show that two such simple extensions can be produced by adjoining different solutions of one irreducible zero-order difference polynomial.

Let P and Q be the algebraically irreducible zero-order difference polynomials annulled by c and d respectively. Considering $\mathfrak{F}$ as an algebraic field, we adjoin to it a solution γ of P and a solution δ of Q, forming the algebraic field $\mathfrak{F}_0$. There is an integer k such that $\mathfrak{F}(\gamma + k\delta)$ is $\mathfrak{F}_0$. Let R be the algebraically irreducible polynomial with coefficients in $\mathfrak{F}$ which is annulled by $\gamma + k\delta$.

Using Theorem IV of [1] in place of the fundamental theorem of algebra, we adjoin to $\mathfrak{H}'$ solutions of the difference polynomial Q equal in number to its degree. Among these solutions there must be a $\bar{d}$ such that the mapping $c \rightarrow \gamma$, $\bar{d} \rightarrow \delta$ carries $\mathfrak{F}(c, \bar{d})$ isomorphically into $\mathfrak{F}_0$. We denote by Σ_1 the reflexive prime difference ideal consisting of all difference polynomials in the unknown y, and with coefficients in $\mathfrak{F}$, which are annulled by $c + k\bar{d}$. Now $\mathfrak{F}(c + k\bar{d})$ must contain $\mathfrak{F}(c)$. Hence $\mathfrak{F}\langle c + k\bar{d}\rangle$ contains $\mathfrak{F}\langle c\rangle$. In other words, the field $\mathfrak{F}_1$ obtained by adjoining to $\mathfrak{F}$ a generic zero of Σ_1 contains a subfield isomorphic to $\mathfrak{H}'$.

---

where each γ_i is the solution of a zero-order difference polynomial with coefficients in $\mathfrak{F}$. It must not be confused with a finite algebraic extension of $\mathfrak{F}$ considered as an algebraic field, that is, with an extension of the form $\mathfrak{F}(\delta_1, \cdots, \delta_r)$, where the δ_i satisfy algebraic equations with coefficients in $\mathfrak{F}$.

[21] A shuffling, as defined in [4], involves taking transforms, forming linear combinations, and factoring difference polynomials into products of transforms.

The difference polynomial P has a solution $\bar{c}$ in an extension of $\mathfrak{F}\langle d\rangle$, which is such that the mapping $\bar{c}\rightarrow\gamma$, $d-\delta$ carries $\mathfrak{F}(\bar{c}, d)$ isomorphically into $\mathfrak{F}_0$. Let Σ_2 be the reflexive prime difference ideal consisting of all difference polynomials in the unknown y, and with coefficients in $\mathfrak{F}$, which are annulled by $\bar{c}+kd$. We see as above that the field $\mathfrak{F}_2$ obtained by adjoining to $\mathfrak{F}$ a generic zero of Σ_2 contains a subfield isomorphic to $\mathfrak{G}'$. Hence $\mathfrak{F}_1$ and $\mathfrak{F}_2$ are incompatible. Since Σ_1 and Σ_2 are essential prime ideals in the decomposition of $\{R\}$, Theorem I is proved.

COROLLARY. *If the difference field $\mathfrak{F}$ is algebraically closed, there is an extension $\mathfrak{K}$ of $\mathfrak{F}$ such that every reflexive prime difference ideal with coefficients in $\mathfrak{F}$ has a generic zero in $\mathfrak{K}$.*

We suppose that the reflexive prime difference ideals with coefficients in $\mathfrak{F}$ have been well-ordered. The proof is then easily carried out by induction. We produce a correspondence between the ideals and certain extensions of $\mathfrak{F}$, such that the extension corresponding to a reflexive prime difference ideal Σ contains a generic zero of Σ, and such that if Σ_1 precedes Σ_2 in the well-ordering, then the extension corresponding to Σ_2 includes that corresponding to Σ_1. Evidently the extension corresponding to the prime ideal which is first in the well-ordering can be constructed. Suppose we have constructed such a set of extensions for all the reflexive prime difference ideals preceding an ideal Λ. Then the union $\mathfrak{G}$ of these extensions is itself an extension of $\mathfrak{F}$. By Theorem I, there is an extension $\mathfrak{H}$ of $\mathfrak{G}$ which contains a generic zero of Λ. We let $\mathfrak{H}$ be the extension of $\mathfrak{F}$ corresponding to Λ. The union of all the extensions of the set is the extension $\mathfrak{K}$ of the corollary.

6. Monadic extensions.

THEOREM II. *If the difference field $\mathfrak{F}$ has a finite monadic extension, then it has a monadic extension $\mathfrak{F}\langle\alpha\rangle$, where α is a solution of a zero-order difference polynomial with coefficients in $\mathfrak{F}$.*

Hence, if $\mathfrak{F}$ is algebraically closed it does not have a finite monadic extension.

Proof. Let $\mathfrak{G}$ be a monadic extension of $\mathfrak{F}$ formed by adjoining to $\mathfrak{F}$ the elements α_i, $i=1,\cdots,n$. The α_i annul some non-zero difference polynomial with coefficients in $\mathfrak{F}$; for otherwise, if we adjoined to $\mathfrak{G}$ a set of elements β_i, $i=1,\cdots,n$ annulling no non-zero difference polynomial with

coefficients in $\mathfrak{G}$, the resulting extension of $\mathfrak{G}$ would contain an extension of $\mathfrak{F}$ isomorphic to $\mathfrak{G}$ and distinct from $\mathfrak{G}$, namely $\mathfrak{F}\langle\beta_1, \cdots, \beta_n\rangle$.

Let $\mathfrak{G}^*$ be the subfield of $\mathfrak{G}$ consisting of all elements of $\mathfrak{G}$ which are algebraic over the inversive extension [22] $\tilde{\mathfrak{F}}$ of $\mathfrak{F}$. We shall show that $\mathfrak{G}$ coincides with $\mathfrak{G}^*$. We assume the contrary and obtain a contradiction.

An element of $\mathfrak{G}$ whose transform is in $\mathfrak{G}^*$ is itself in $\mathfrak{G}^*$. Hence, by our assumption, $\mathfrak{G}$ contains elements no transform of which is in $\mathfrak{G}^*$. Then $\mathfrak{G}$ must be a monadic extension of $\mathfrak{G}^*$ formed by adjoining the α_i. Let Σ be the reflexive prime difference ideal consisting of all polynomials in $y_1, \cdots, y_n$, with coefficients in $\mathfrak{G}^*$, which vanish when the y_i are replaced by the corresponding α_i. Σ is not empty. Since there is an α_i no transform of which is algebraic over $\mathfrak{G}^*$, Σ either has a set of parametric indeterminates which is not empty, or is of effective order different from zero.

Let Λ denote $\{\Sigma; \mathfrak{G}\}$. By Lemma III, Λ is a reflexive prime difference ideal which, like Σ, either has parametric indeterminates, or is not of effective order zero. Hence a generic zero $\beta_1, \cdots, \beta_n$ of Λ is not contained in $\mathfrak{G}$. By the corollary to Lemma IV, Λ contains no polynomial with coefficients in $\mathfrak{G}^*$ except the polynomials of Σ. Hence the extension $\mathfrak{G}\langle\beta_1, \cdots, \beta_n\rangle$ contains a subfield $G^*\langle\beta_1, \cdots, \beta_n\rangle$ isomorphic to and distinct from $\mathfrak{G}$. This contradicts the fact that $\mathfrak{G}$ is a monadic extension of $\mathfrak{G}^*$. Hence $\mathfrak{G} = \mathfrak{G}^*$.

From what we have just shown, every α_i has a transform α_i^* algebraic over $\mathfrak{F}$. $\mathfrak{F}\langle\alpha_1^*, \cdots, \alpha_n^*\rangle$ is a monadic extension of $\mathfrak{F}$. For first of all, there is some α_i, and hence some α_i^*, no transform of which is in $\mathfrak{F}$. Secondly, if there existed an extension $\mathfrak{K}$ of $\mathfrak{F}$ such that $\mathfrak{F}\langle\alpha_1^*, \cdots, \alpha_n^*\rangle$ had two distinct isomorphisms into $\mathfrak{K}$, then $\mathfrak{F}\langle\alpha_1, \cdots, \alpha_n\rangle$ would have two distinct isomorphisms into the inversive extension of $\mathfrak{K}$, which is impossible.

As we showed in the proof of Theorem I, a finite zero-order extension of a difference field is simple. Hence there exists an element α such that $\mathfrak{F}\langle\alpha\rangle = \mathfrak{F}\langle\alpha_1^*, \cdots, \alpha_n^*\rangle$. α is the solution of a zero-order difference polynomial with coefficients in $\mathfrak{F}$. This proves Theorem II.

---

[22] An inversive extension of a difference field is an extension which, if it contains an element h, contains an element g whose transform is h. In particular, we mean by *the* inversive extension $\tilde{\mathfrak{F}}$ of a difference field $\mathfrak{F}$, that unique inversive extension all of whose elements have transforms in $\mathfrak{F}$. It existence is shown in [2]. The inversive extension $\tilde{\mathfrak{G}}$ of $\mathfrak{G}$ contains $\tilde{\mathfrak{F}}$. Those elements of $\tilde{\mathfrak{G}}$ which are elements of $\mathfrak{G}$ and which are algebraic over $\tilde{\mathfrak{F}}$ are the elements of $\mathfrak{G}$ which we call algebraic over the inversive extension of $\mathfrak{F}$.

COROLLARY. *If $\mathfrak{G}$ is a finite monadic extension of the difference field $\mathfrak{F}$, then every element of $\mathfrak{G}$ is algebraic over the inversive extension of $\mathfrak{F}$.*

Proof. This is equivalent to the statement $\mathfrak{G} = \mathfrak{G}^*$, proved in the preceding demonstration.

An examination of the proof of Theorem II will show that we have made no use of the fact that the monadic extension $\mathfrak{G}$ has no automorphisms onto itself except the identity. We have used only the fact that $\mathfrak{G}$ cannot be isomorphic to two distinct subfields of an extension of $\mathfrak{F}$. Hence, *if $\mathfrak{G}$ is a finite extension of $\mathfrak{F}$ which has this property, then there is a simple zero-order extension of $\mathfrak{F}$ which contains a transform of every element of $\mathfrak{G}$.*

Conversely, if $\mathfrak{G}' = \mathfrak{F}\langle\alpha\rangle$ is a simple zero-order extension of $\mathfrak{F}$, there is an extension $\mathfrak{G}$ of $\mathfrak{G}'$ which has the property just stated. For let P be the algebraically irreducible difference polynomial whose solution is α. P is of zero order in an indeterminate y, and of degree $n \geqq 1$ in y_0. Using Theorem IV of [1] in place of the fundamental theorem of algebra, we may adjoin $n - 1$ additional solutions of P to $\mathfrak{G}'$. The resulting field is $\mathfrak{G}$.

This property of zero-order extensions is in no way surprising; we find quite the same situation in algebraic fields and differential fields. What is anomalous is that some of these extensions have no automorphisms except the identity, i. e. are monadic. We have found no other way of characterizing those zero-order extensions which are monadic. At the end of this paper we give a conjecture which, if correct, gives a somewhat narrower class of extensions which includes the finite monadic extensions.

7. Infinite extensions. We obtain conclusions somewhat weaker than those of the preceding sections when we deal with countably infinite extensions of a difference field which are incompatible or monadic. To study these, we must first consider the following situation.

Let $\mathfrak{H}$ be an extension of the difference field $\mathfrak{H}^*$, such that every element of $\mathfrak{H}$ not included in $\mathfrak{H}^*$ is transcendental over $\mathfrak{H}^*$. Let a set of elements $\alpha_1, \cdots, \alpha_n, \alpha_{n+1}$ be adjoined to $\mathfrak{H}$. Let Σ_n denote the reflexive prime difference ideal consisting of all polynomials in $y_1, \cdots, y_n$, with coefficients in $\mathfrak{H}^*$, which vanish when these y_i are replaced by the corresponding α_i. Let Σ_{n+1} consist of all polynomials in $y_1, \cdots, y_n, y_{n+1}$, with coefficients in $\mathfrak{H}^*$, which vanish when these y_i are replaced by the corresponding α_i. Of course Σ_n consist of those polynomials of Σ_{n+1} which are free of y_{n+1}. Let $\Lambda_n = \{\Sigma_n ; \mathfrak{H}\}$,

and let $\Lambda_{n+1} = \{\Sigma_{n+1}; \mathfrak{H}\}$. We shall show that Λ_n *consists of those polynomials of* Λ_{n+1} *which are free of* y_{n+1}.

We denote by Σ_{nk} and $\Sigma_{n+1\,k}$, $k = 0, 1, \cdots$, the polynomial ideals consisting of those polynomials of Σ_n and Σ_{n+1}, respectively, which contain no transforms of any y_i of order exceeding k. Let $\Lambda_{nk} = \{\Sigma_{nk}; \mathfrak{H}\}_0$, and let $\Lambda_{n+1\,k} = \{\Sigma_{n+1\,k}; \mathfrak{H}\}_0$. We see from the proof of Lemma III that Λ_n consists of those difference polynomials a transform of which is in some Λ_{nk}, and that Λ_{n+1} consists of those difference polynomials a transform of which is in some $\Lambda_{n+1\,k}$.

Σ_{nk} consists of all polynomials of $\Sigma_{n+1\,k}$ which are free of transforms of y_{n+1}. We choose a complete set T_{nk} of parametric indeterminates for Σ_{nk}. (T_{nk} may be empty, or may consist—when Σ_{nk} is the ideal $[0]$—of all the indeterminates of Σ_{nk}.) Since $\Sigma_{n+1\,k}$ contains no non-zero polynomial in the indeterminates of T_{nk}, we may select a complete set $T_{n+1\,k}$ of parametric indeterminates for $\Sigma_{n+1\,k}$ which includes T_{nk}. T_{nk} and $T_{n+1\,k}$ are also sets of parametric indeterminates for Λ_{nk} and $\Lambda_{n+1\,k}$ respectively, by Lemma II.

Let Π be the prime polynomial ideal which consists of those polynomials of $\Lambda_{n+1\,k}$ which are free of y_{n+1} and its transforms. Then $\Pi \supset \Lambda_{nk}$. Since Π contains no polynomial in the indeterminates of T_{nk}, it has at least as large a set of parametric indeterminates as Λ_{nk}. Hence Λ_{nk} cannot be a proper subset of Π. Then $\Pi = \Lambda_{nk}$. In other words Λ_{nk} consists of those polynomials of $\Lambda_{n+1\,k}$ which are free of y_{n+1} and its transforms.

Let P be a difference polynomial in Λ_{n+1} which does not involve y_{n+1}. Then for some k, a transform of P is in $\Lambda_{n+1\,k}$, and therefore in Λ_{nk}. Thus P is in Λ_n, which proves our statement.

Let $\gamma_1, \cdots, \gamma_{n+1}$ be a generic zero of Λ_{n+1}. By the result just proved, $\gamma_1, \cdots, \gamma_n$ is a generic zero of Λ_n. Conversely, let $\alpha_1, \cdots, \alpha_n$ be a generic zero of Λ_n. Since $\mathfrak{H}' = \mathfrak{H}\langle\alpha_1, \cdots, \alpha_n\rangle$ is isomorphic to $\mathfrak{H}\langle\gamma_1, \cdots, \gamma_n\rangle$, we can find an extension of $\mathfrak{H}'$ which is isomorphic to $\mathfrak{H}\langle\gamma_1, \cdots, \gamma_n, \gamma_{n+1}\rangle$. Let α_{n+1} be the element corresponding to γ_{n+1} in this isomorphism. Then $\alpha_1, \cdots, \alpha_n, \alpha_{n+1}$ is a generic zero of Λ_{n+1}. Hence *any generic zero of* Λ_n *can be extended to a generic zero of* Λ_{n+1}. It is in this form that our result will be used in studying countably infinite extensions.

THEOREM III. *Let a difference field* $\mathfrak{F}$ *have the at most countably infinite extensions* $\mathfrak{G}$ *and* $\mathfrak{H}$ *which are incompatible. Then* $\mathfrak{F}$ *has incompatible extensions* $\mathfrak{G}^*$ *and* $\mathfrak{H}^*$, *at most countably infinite, every element of which is algebraic over* $\mathfrak{F}$.

Proof. Let $\mathfrak{G}^*$ and $\mathfrak{H}^*$ consist of those elements of $\mathfrak{G}$ and $\mathfrak{H}$, respectively, which are algebraic over $\mathfrak{F}$. We shall assume that $\mathfrak{G}$ and $\mathfrak{H}^*$ are not incompatible, and show that this implies that $\mathfrak{G}$ and $\mathfrak{H}$ are not incompatible.

By our assumption, there is an extension $\mathfrak{H}'$ of $\mathfrak{H}^*$ which contains a subfield isomorphic to $\mathfrak{G}$. This subfield must be formed by adjoining to $\mathfrak{F}$ a set of elements α_i, $i = 1, 2, \cdots$. There is no loss of generality in assuming that there are an infinite number of α_i, since they need not all be distinct.

For each n, $n = 1, 2, \cdots$, let Σ_n denote the reflexive prime difference ideal of polynomials of $\mathfrak{H}^*\{y_1, \cdots, y_n\}$ which vanish when each y_i is replaced by α_i. Let Λ_n be the reflexive prime difference ideal $\{\Sigma_n; \mathfrak{H}\}$. We shall adjoin a set of elements β_i, $i = 1, 2, \cdots$, to $\mathfrak{H}$ such that, for each $n \geqq 1$, $\beta_1, \cdots, \beta_n$ is a generic zero of Λ_n. If Λ_n is $[0]$, this is to mean that $\beta_1, \cdots, \beta_n$ annul no non-zero difference polynomial with coefficients in $\mathfrak{H}$.

For β_1 we choose a generic zero of Λ_1. Suppose we have adjoined $\beta_1, \cdots, \beta_k$. Then $\beta_1, \cdots, \beta_k$ is a generic zero of Λ_k. Now Σ_k consists of those polynomials of Σ_{k+1} which are free of y_{k+1}. By the result obtained at the beginning of this section, we may adjoin an element to the set $\beta_1, \cdots, \beta_k$ to obtain a generic zero of Λ_{k+1}. Let β_{k+1} be this element. By induction we construct the set β_i, $i = 1, 2, \cdots$.

For any n, $\beta_1, \cdots, \beta_n$ annul no polynomial of $\mathfrak{F}\{y_1, \cdots, y_n\}$, except those annulled by $\alpha_1, \cdots, \alpha_n$. For, by the corollary to Lemma IV, Λ_n contains no polynomial with coefficients in $\mathfrak{H}^*$ except the polynomials of Σ_n. Hence Λ_n contains no polynomial with coefficients in $\mathfrak{F}$, except the polynomials of Σ_n with coefficients in $\mathfrak{F}$. It follows that $\mathfrak{H}\langle\beta_1, \beta_2, \cdots\rangle$ contains a subfield isomorphic to $\mathfrak{G}$, namely $\mathfrak{F}\langle\beta_1, \beta_2, \cdots\rangle$. Then $\mathfrak{G}$ and $\mathfrak{H}$ are not incompatible. Since we have obtained a result which contradicts the hypothesis of Theorem III, it follows that $\mathfrak{G}$ and $\mathfrak{H}^*$ are incompatible.

We can find a countable set of elements $\gamma_1, \gamma_2, \cdots$, such that the field $\mathfrak{F}(\gamma_1, \gamma_2, \cdots)$, formed by adjoining the γ_i to $\mathfrak{F}$ considered as an algebraic field, is $\mathfrak{H}$. We need merely take for the γ_i all transforms of the elements which produce $\mathfrak{H}$ when they are adjoined to $\mathfrak{F}$, considered as a difference field. It follows that when $\mathfrak{H}$ is considered as a linear vector space with $\mathfrak{F}$ for coefficient field, its dimension is at most $\aleph_0$. Hence $\mathfrak{H}^*$ is a linear vector space over $\mathfrak{F}$ of dimension at most $\aleph_0$. Therefore $\mathfrak{H}^*$ is an at most countably infinite extension of $\mathfrak{F}$, considered as a difference field. By a repetition of the demonstration above, we see that $\mathfrak{H}^*$ is incompatible with $\mathfrak{G}^*$. This proves Theorem III.

THEOREM IV. *Let a difference field $\mathfrak{F}$ have a monadic extension $\mathfrak{G}$ which is an at most countably infinite extension of $\mathfrak{F}$. Then every element of $\mathfrak{G}$ has a transform which is algebraic over $\mathfrak{F}$.*

Proof. Let $\mathfrak{G}^*$ be the set of elements of $\mathfrak{G}$ which have transforms algebraic over $\mathfrak{F}$. We shall assume that $\mathfrak{G}$ contains elements which are not in $\mathfrak{G}^*$ and obtain a contradiction. $\mathfrak{G}$ can be obtained by adjoining a countable set of elements $\alpha_1, \alpha_2, \cdots$, to $\mathfrak{F}$. Of these elements, at least one is not in $\mathfrak{G}^*$ because of our assumption. Changing the subscripts if necessary, we arrange matters so that α_1 is not in $\mathfrak{G}^*$. Then α_1 annuls no difference polynomial of effective order zero and with coefficients in $\mathfrak{G}^*$. For otherwise, a transform of α_1 would be algebraic over $\mathfrak{F}$, and, α_1 would be in $\mathfrak{G}^*$.

Let Σ_n denote the reflexive prime difference ideal consisting of those polynomials of $\mathfrak{G}^*\{y_1, \cdots, y_n\}$ which vanish when the y_i are replaced by the α_i. Let Λ_n be the reflexive prime difference ideal $\{\Sigma_n; \mathfrak{G}\}$.

We shall adjoin a set of elements $\beta_1, \beta_2, \cdots$ to $\mathfrak{G}$ such that, for each $n \geqq 1$, $\beta_1, \cdots, \beta_n$ is a generic zero of Λ_n. For β_1 we choose a generic zero of Λ_1. Suppose we have already adjoined $\beta_1, \cdots, \beta_k$. Then $\beta_1, \cdots, \beta_k$ is a generic zero of Λ_k. By the result obtained at the begnning of this section, we may adjoin an element β_{k+1} to the set $\beta_1, \cdots, \beta_k$ so as to obtain a generic zero of Λ_{k+1}. By induction we construct the set β_i, $i = 1, 2, \cdots$.

By the provision above concerning α_1, Σ_1 is either $[0]$ or of effective order different from zero. Hence Λ_1 is $[0]$ or of effective order different from zero. Therefore its generic zero β_1 is not in the coefficient field $\mathfrak{G}$.

For each n, Λ_n contains no difference polynomial with coefficients in $\mathfrak{F}$, except the polynomials of Σ_n which have coefficients in $\mathfrak{F}$. Hence $\mathfrak{F}\langle\beta_1, \beta_2, \cdots\rangle$ is isomorphic to $\mathfrak{G}$. Then $\mathfrak{G}\langle\beta_1, \beta_2, \cdots\rangle$ contains two subfields isomorphic to $\mathfrak{G}$, namely $\mathfrak{G}$ itself and $\mathfrak{F}\langle\beta_1, \beta_2, \cdots\rangle$. These subfields are distinct, because β_1 is not in $\mathfrak{G}$. Therefore $\mathfrak{G}$ is not a monadic extension of $\mathfrak{F}$, and we have obtained a contradiction. This proves Theorem IV.[23]

8. Conclusion. We have seen that if a difference field $\mathfrak{F}$ permits finite incompatible extensions, or a finite monadic extension, then there is a single algebraically irreducible difference polynomial A of zero order in its indeter-

[23] As noted after the proof of Theorem II, we have actually proved a theorem concerning a class of extensions wider than the class of monadic extensions. For it is only necessary to assume that $\mathfrak{G}$ is such that no extension of $\mathfrak{F}$ contains two distinct subfields isomorphic to $\mathfrak{G}$.

minate y, such that a monadic extension of $\mathfrak{F}$, or incompatible extensions of $\mathfrak{F}$, can be obtained by adjoining solutions of A.

Now for a single difference polynomial of zero order the two cases of incompatible extensions and of a monadic extension represent opposite extremes of a range of possibilities. Let us consider such a polynomial A, algebraically irreducible, and of degree $n > 1$ in y. We suppose that $\{A\}$ is the intersection of two or more reflexive prime difference ideals $\Pi_1, \cdots, \Pi_r$, whose manifolds we denote by $\mathfrak{M}_1, \cdots, \mathfrak{M}_r$ respectively.

We show as in the proof of Theorem I that there exist extensions of $\mathfrak{F}$ containing n solutions of A. In fact any difference field containing $k < n$ solutions of A can be extended to a field containing n solutions of A. Not all such extensions of $\mathfrak{F}$ need be isomorphic, as we see from Example 1 above. It may be that for a particular polynomial A, the solutions in any one such extension all lie in the same manifold $\mathfrak{M}_i$ (where i depends on the extension). In such a case all the manifolds are evidently incompatible. Examples 1-3 all illustrate this. If, on the other hand, every such extension of $\mathfrak{F}$ contains one solution in each $\mathfrak{M}_i$ (so that we must have $r = n$), then each $\mathfrak{M}_i$ is a monadic manifold. Example 4 illustrates this.

For an example of a distribution of the solutions among the manifolds which is intermediate between these extremes, and which results neither in monadic nor in incompatible extensions, we consider the field $\mathfrak{T}$ which is formed by adjoining a transformally transcendental element u to a difference field $\mathfrak{F}$.

We consider $\mathfrak{T}$ momentarily as an alegbraic field and adjoin a sequence of elements $\alpha_0, \alpha_1, \cdots$, such that $\alpha_i^2 = u_i$, $i = 0, 1, \cdots$. Evidently $\mathfrak{T}(\alpha_0, \alpha_1, \cdots)$ becomes a difference field $(\epsilon)\mathfrak{T}$ if we define the transform of α_i to be $\epsilon_i\alpha_{i+1}$, $i = 0, 1, \cdots$, where each ϵ_i is either $+1$ or -1. All the difference fields $(\epsilon)\mathfrak{T}$ are isomorphic. There is no other way of extending the endomorphism of $\mathfrak{T}$ to $\mathfrak{T}(\alpha_0, \alpha_1, \cdots)$.

Let A be the algebraically irreducible polynomial in y, with coefficients in $\mathfrak{T}$, whose solutions are $\alpha_0 + \alpha_1$, $\alpha_0 - \alpha_1$, $-\alpha_0 + \alpha_1$, $-\alpha_0 - \alpha_1$. We consider A is a difference polynomial. Let γ be any solution of A. Since α_i can be expressed rationally in terms of $\alpha_i + \alpha_{i+1}$ or any of its conjugates, $T\langle\gamma\rangle$ is isomorphic with one of the fields $(\epsilon)\mathfrak{T}$ and therefore contains four solutions of A. Since the fields $(\epsilon)\mathfrak{T}$ are isomorphic, all fields formed by adjoining solutions of A to $\mathfrak{T}$ are isomorphic. Hence the four solutions must be distributed in a fixed manner among the essential irreducible manifolds of A, and there must be at least one of them in each of these manifolds. We

shall show that there are actually two essential irreducible manifolds of A, and that two of the four solutions lie in each of them.

For convenience, let γ be a solution in one of the fields $(\epsilon)\mathfrak{T}$. Then γ_0 is a conjugate of $\alpha_0 + \alpha_1$. We form the algebraic field $\mathfrak{T}_1 = \mathfrak{T}(\gamma_0)$ which must contain α_0 and α_1. α_2 is of second degree over $\mathfrak{T}_1$. Hence γ_1 is of second degree over $\mathfrak{T}_1$. Then the transform A_1 of A factors in $\mathfrak{T}_1$ into two irreducible polynomials of second degree. Let $\mathfrak{T}_k$ denote $\mathfrak{T}(\alpha_0, \alpha_1, \cdots, \alpha_{k-1})$. It follows as with A_1 that the h-th transform A_k of A factors in $\mathfrak{T}_k$ into two algebraically irreducible polynomials of second degree. Theorem IV of [1] now shows that A has two irreducible manifolds $\mathfrak{M}_1$ and $\mathfrak{M}_2$.

From our previous discussion we know that if we adjoin a solution of either $\mathfrak{M}_1$ or $\mathfrak{M}_2$ to $\mathfrak{T}$, we obtain a field isomorphic to any $(\epsilon)\mathfrak{T}$, and containing a solution γ' of $\mathfrak{M}_1$, and a solution γ'' of $\mathfrak{M}_2$. But the mapping which carries each α_i into $-\alpha_i$, or the corresponding mapping in a field isomorphic to $(\epsilon)\mathfrak{T}$, is easily seen to be an automorphism of the difference field $(\epsilon)\mathfrak{T}$. It carries γ' into $-\gamma'$, and γ'' into $-\gamma''$. Hence $-\gamma'$ and $-\gamma''$ are solution of $\mathfrak{M}_1$ and $\mathfrak{M}_2$ respectively. This proves that there are always two solutions in each manifold. Evidently $\mathfrak{M}_1$ and $\mathfrak{M}_2$ are not incompatible, nor does either furnish a monadic extension of $\mathfrak{T}$.

We may duplicate this example in the field $\mathfrak{R}$ of rational functions of x with complex coefficients, the transform of a function $f(x)$ being defined as $f(x+1)$. We simply replace the u_k in the preceding example by $x + k$.

It would be interesting to know whether the fields $\mathfrak{R}$ or $\mathfrak{T}$ have finite monadic or incompatible extensions (other than those obtainable from incompatible or monadic extensions of $\mathfrak{F}$ by adjoining x or u). It would follow that they do not if we could establish the following *conjecture*: If an essential irreducible manifold $\mathfrak{M}$ of an algebraically irreducible zero order difference polynomial A with coefficients in a difference field $\mathfrak{F}$ is monadic, or if it is incompatible with another manifold of A, then a solution of $\mathfrak{M}$ is contained in an extension of $\mathfrak{F}$ which, considered as an algebraic field, is a finite extension. In other terms, we conjecture that the characteristic set of A which is annulled by $\mathfrak{M}$ terminates in a polynomial which is of first degree in the highest transform of the indeterminate present in it.

RUTGERS UNIVERSITY.

BIBLIOGRAPHY.

[1] R. M. Cohn, "Manifolds of difference polynomials," *Transactions of the American Mathematical Society*, vol. 64 (1948), pp. 133-172.

[2] ———, "Inversive difference fields," *Bulletin of the American Mathematical Society*, vol. 55 (1949), pp. 597-603.

[3] J. L. Doob and J. F. Ritt, "Systems of algebraic difference equations," *American Journal of Mathematics*, vol. 55 (1933), pp. 505-514.

[4] H. W. Raudenbush and J. F. Ritt, "Ideal theory and algebraic difference equations," *Transactions of the American Mathematical Society*, vol. 46 (1939), pp. 445-452.

[5] J. F. Ritt, "Complete difference ideals," *American Journal of Mathematics*, vol. 63 (1941), pp. 681-690.

[6] ———, "Differential algebra," *Colloquium Publications of the American Mathematical Society*, vol. 33.

ON THE COHOMOLOGY THEORY FOR ASSOCIATIVE ALGEBRAS.* [1]

By I. H. ROSE.

1. Introduction. Cohomology theory in general (cf. [1]) concerns itself with the following situation, which arose originally in topology:

$C^0, C^1, \cdots, C^n, \cdots$ is a sequence of abelian groups, $\delta^0, \delta^1, \cdots, \delta^n, \cdots$ a sequence of homomorphisms such that δ^n maps C^n into C^{n+1} and such that $\delta^{n+1}\delta^n = 0$.

In this situation the elements of C^n are called n-dimensional cochains; the kernel of δ^n is denoted Z^n and its elements are called n-dimensional cocycles; the image of δ^n is denoted B^{n+1} and its elements are called $(n+1)$-dimensional coboundaries; B^0 is defined to be the zero element of C^0. Since $\delta^{n+1}\delta^n = 0$, it follows that $B^n \subset Z^n$. We may therefore define the "cohomology" group $H^n = Z^n/B^n$; two n-dimensional cocycles are called cohomologous if their difference is a coboundary. The symbol δ is used to represent any of the homomorphisms δ^n, since this leads to no ambiguity.

The cohomology theory for associative algebras specializes the general situation as follows.

Let F be a field over which are defined an associative algebra A and a vector space P. Suppose further that P is a two-sided A-module, i. e. for each $a \in A$, $p \in P$ there are defined elements $a \cdot p$, $p \cdot a \in P$ which are bilinear functions of a and p such that

$$a_1 \cdot (a_2 \cdot p) = a_1 a_2 \cdot p, \qquad (p \cdot a_1) \cdot a_2 = p \cdot a_1 a_2, \qquad a_1 \cdot (p \cdot a_2) = (a_1 \cdot p) \cdot a_2.$$

Now for $n > 0$ define $C^n = C^n(A, P)$ as the vector space of all n-linear functions of n variables mapping A into P, and define $C^0 = C^0(A, P) = P$. Furthermore, for $n > 0$ and $f \in C^n(A, P)$ define δf such that:

$$\begin{aligned}\delta f(a_1, \cdots, a_{n+1}) &= a_1 \cdot f(a_2, \cdots, a_{n+1}) - f(a_1 a_2, a_3, \cdots, a_{n+1}) \\ &+ \sum_{i=2}^{n-1} (-1)^i f(a_1, \cdots, a_{i-1}, a_i a_{i+1}, a_{i+2}, \cdots, a_{n+1}) \\ &+ (-1)^n f(a_1, \cdots, a_{n-1}, a_n a_{n+1}) - (-1)^n f(a_1, \cdots, a_n) \cdot a_{n+1}.\end{aligned}$$

Finally, for $n = 0$, define $\delta p(a) = a \cdot p - p \cdot a$.

* Received May 16, 1951.

[1] This paper is based on a portion of the author's doctoral dissertation at Harvard University.

It is then easily proved (see [4], p. 60) that δ is a vector space homomorphism such that $\delta\delta = 0$, so that we have here a special case of the preceding general situation. In this case we denote Z^n by $Z^n(A, P)$, B^n by $B^n(A, P)$ etc.

This paper is concerned with several problems suggested by Hochschild. We consider first a problem proposed in [4]. After introducing for a given algebra the condition C_m: "All m-dimensional cohomology groups vanish," the statement is made (p. 58) "Theorem 3.1 implies that C_{m+1} is a consequence of C_m for $m \geq 1$. But it is an an open question whether or not C_m and C_{m+1} are equivalent."

We denote by K_m, $m = 0, 1, 2, \cdots$ the class of all algebras over F whose m-dimensional cohomology groups are all zero, or in other words cocycles are all coboundaries. (It is easily proved that K_0 is the null class.) One may then paraphrase the question concerning the condition C_m as follows: For $m \geq 0$ we have $K_m \subset K_{m+1}$; do we actually have $K_m = K_{m+1}$?

For $m = 0, 1, 2$ Hochschild has proved that the answer to this question is in the negative. The proof for $m = 0$ follows from his demonstration that the algebras in K_1 are a well-known non-null set, namely the set of algebras separable over F ([4], Theorem 4.1). For $m = 1$ an algebra (which we denote H') is produced such that $H' \varepsilon K_2 - K_1$ ([4], section 9). The case $m = 2$ is disposed of by first proving that adjoining an identity to an algebra does not affect its K-class ([5], section 2); then, letting H be the algebra formed by adjoining an identity to H', it is proved that the Kronecker product $H \times H \varepsilon K_3 - K_2$ ([6], Theorem 9.2).

Immediately following Hochschild's proof of this last result he states a conjecture which we shall call Conjecture 1.

CONJECTURE 1. *The n-fold Kronecker product $H \times H \times \cdots \times H$ does not belong to K_n.*

We prove Conjecture 1 in Prop. 6.1. There easily follows (Theorem 6.1) the answer to the first question raised, namely: $K_m \neq K_{m+1}$; for although the n-fold product $H \times H \times \cdots \times H$ does not belong to K_n, it does, as a consequence of Theorem 5.1, belong to K_{n+1}.

Conjecture 1 is followed in [6] by another conjecture:

CONJECTURE 2. *If $A \varepsilon K_p - K_{p-1}$ and $B \varepsilon K_q - K_{q-1}$, then the Kronecker product $A \times B \varepsilon K_{p+q-1} - K_{p+q-2}$.*

We shall consider in this paper the following two conjectures which together imply Conjecture 2.

CONJECTURE 2a. $A \varepsilon K_p, B \varepsilon K_q \Longrightarrow A \times B \varepsilon K_{p+q-1}$.

CONJECTURE 2b. $A \not\varepsilon K_p, \quad B \not\varepsilon K_q \Longrightarrow A \times B \not\varepsilon K_{p+q}$.

A special case of Conjecture 2a, namely that in which A and B have identities and $q = 2$, is a consequence of the following theorem in [6]:

THEOREM 9.1. *Let A and B be algebras with identity elements, and suppose that all two-dimensional cohomology groups of B are zero. Then for every $A \times B$ module Q and $n \geq 1$, we have*

$$H^n[A, Z^0(B, C_1(A \times B, Q))] \approx H^{n+1}(A \times B, Q).$$

In [6] the proof of Theorem 9.1 is followed by another conjecture:

CONJECTURE 3. *If A and B have identities, and $B \varepsilon K_{p+1}$, then*

$$H^n[A, Z^0(B, C^p(A \times B, Q))] \approx H^{n+p}(A \times B, Q).$$

We prove Conjecture 3 in Theorem 4.1. There immediately follows (Theorem 5.1) that Conjecture 2a is true for algebras with identities. (Conjectures 2b and 2 are definitely not true for algebras A, B without identities; a counterexample will be exhibited in a subsequent paper on the classification of algebras by means of cohomology theory). For the case $p = 0$, A with an identity we prove Conjecture 2b in Cor. 5.1, while in the case $p = 1$, A and B with identities, a proof is given in Theorem 5.2. It is likely that Conjecture 2b is true for algebras A, B with identities, but a proof covering all cases remains to be found.

2. Notations and conventions. Except where otherwise indicated, the following notation and conventions will be assumed.

F: A field over which all algebras in the sequel are defined; all algebras in the sequel have identities.

$A \times B$: The Kronecker product of algebras A, B. If $1_\alpha, 1_\beta$ are the identities of A, B respectively, we identify $a \times 1_\beta$ with $a \varepsilon A$ and $1_\alpha \times b$ with $b \varepsilon B$ in situations where these identifications lead to no ambiguity; also, we denote $1_\alpha \times 1_\beta$ by 1.

S_i: The element $a_i \times b_i \varepsilon A \times B$.

${}_ra_s$: The sequence $a_r, a_{r+1}, \cdots, a_s$, $r \leq s$.

L_p: The class of algebras $K_{p+1} - K_p$, $p = 0, 1, 2, \cdots$; note that $L_0 = K_1$.

3. The D-modules M_n and M'_n. $(n = 0, 1, 2, \cdots)$. Let D be an algebra, M a D-module. We shall find it useful to make the vector space $C^n(D, M)$ into a D-module in two ways.

(I) *For* $n > 0$, $C^n(D, M)$ *is made into a D-module* M_n *as follows: Let* $d, {}_1d_n \varepsilon D$, $g \varepsilon C^n(D, M)$. *Then we define* $d \cdot g$ *and* $g \cdot d$ *such that*

$$\{d \cdot g\}({}_1d_n) = d \cdot g({}_1d_n) \text{ and } \{g \cdot d\}({}_1d_n) = d \cdot g({}_1d_n) - \delta g(d, {}_1d_n).$$

(II) *For* $n > 0$, $C^n(D, M)$ *is made into a D-module* M'_n *as follows: Let* $d, {}_1d_n \varepsilon D$, $g \varepsilon C^n(D, M)$. *Then we define* $d \cdot g$ *and* $g \cdot d$ *such that*

$$\{d \cdot g\}({}_1d_n) = (-1)^n \delta g({}_1d_n, d) + g({}_1d_n) \cdot d \text{ and } \{g \cdot d\}({}_1d_n) = g({}_1d_n) \cdot d.$$

The definitions are completed by setting $M_0 = M'_0 = M$; to verify that M_n, M'_n actually are D-modules involves only straightforward computation.

The usefulness of these modules lies in the following two propositions.

PROPOSITION 3.1. *Given* $f \varepsilon C^{m+n}(D, M)$, $m \geq 0$, $n > 0$, *define* $\bar{f} \varepsilon C^m(D, M_n)$ *such that* $\bar{f}({}_1d_n) = f({}_1d_n)'$ *for* $m = 0$, $\{\bar{f}({}_1d_m)\}({}_{m+1}d_{m+n}) = f({}_1d_{m+n})$ *for* $m > 0$. *Then* $\{\delta \bar{f}({}_1d_{m+1})\}({}_{m+2}d_{m+n+1}) = \delta f({}_1d_{m+n+1})$.

Proof. Note first that for $m > 0$,

$$\begin{aligned}
&\{\bar{f}({}_1d_m) \cdot d_{m+1}\}({}_{m+2}d_{m+n+1}) \\
&\qquad = d_{m+1} \cdot \{\bar{f}({}_1d_m)\}({}_{m+2}d_{m+n+1}) - \{\delta[\bar{f}({}_1d_m)]\}({}_{m+1}d_{m+n+1}) \\
&\qquad = d_{m+1} \cdot \{\bar{f}({}_1d_m)\}({}_{m+2}d_{m+n+1}) - d_{m+1} \cdot \{\bar{f}({}_1d_m)\}({}_{m+2}d_{m+n+1}) \\
&\qquad\quad + f({}_1d_m, d_{m+1}d_{m+2}, {}_{m+3}d_{m+n+1}) + \cdots + (-1)^n f({}_1d_{m+n}) \cdot d_{m+n+1}.
\end{aligned}$$

Therefore, for $m > 0$,

$$\begin{aligned}
&\{\delta\bar{f}({}_1d_{m+1})\}({}_{m+2}d_{m+n+1}) \\
&= d_1 \cdot f({}_2d_{m+n+1}) - f(d_1d_2, {}_3d_{m+n+1}) + \cdots + (-1)^m f({}_1d_{m-1}, d_md_{m+1}, {}_{m+2}d_{m+n+1}) \\
&\quad - (-1)^m \{\bar{f}({}_1d_m) \cdot d_{m+1}\}({}_{m+2}d_{m+n+1}) = \delta f({}_1d_{m+n+1}), \qquad \text{q. e. d.}
\end{aligned}$$

For $m = 0$ the result follows immediately from (I).

PROPOSITION 3.2. *Given* $f \varepsilon C^{m+n}(D, M)$, $m \geq 0$, $n > 0$, *define* $\bar{f} \varepsilon C^m(D, M'_n)$ *such that* $\bar{f}({}_1d_n) = (-1)^n f({}_1d_n)$ *for* $m = 0$, $\{\bar{f}({}_{n+1}d_{n+m})\}({}_1d_n) = (-1)^n f({}_1d_{m+n})$ *for* $m > 0$.

Then $\{\delta\bar{f}({}_{n+1}d_{m+n+1})\}({}_1d_n) = \delta f({}_1d_{m+n+1})$.

Proof. Note first that for $m > 0$,

$$\begin{aligned}
&\{d_{n+1} \cdot \bar{f}({}_{n+2}d_{n+m+1})\}({}_1d_n) \\
&\qquad = (-1)^n \{\delta[\bar{f}({}_{n+2}d_{n+m+1})]\}({}_1d_{n+1}) + \{\bar{f}({}_{n+2}d_{n+m+1})\}({}_1d_n) \cdot d_{n+1} \\
&\qquad = (-1)^n [d_1 \cdot [\{\bar{f}({}_{n+2}d_{n+m+1})\}({}_2d_{n+1})] - \cdots \\
&\qquad\qquad + (-1)^n \{\bar{f}({}_{n+2}d_{n+m+1})\}({}_1d_{n-1}, d_nd_{n+1})] \\
&\qquad = d_1 \cdot f({}_2d_{m+n+1}) - \cdots + (-1)^n f({}_1d_{n-1}, d_nd_{n+1}, {}_{n+2}d_{m+n+1}).
\end{aligned}$$

Therefore, for $m > 0$,

$$\{\delta\bar{f}({}_{n+1}d_{m+n+1})\}({}_1d_n)$$
$$= \{d_{n+1}\cdot\bar{f}({}_{n+2}d_{n+m+1})\}({}_1d_n) - (-1)^n f({}_1d_n, d_{n+1}d_{n+2}, d_{n+3}d_{n+m+1}) + \cdots$$
$$+ (-1)^n(-1)^{m+1}f({}_1d_{n+m})\cdot d_{n+m+1} = \delta f({}_1d_{m+n+1}), \qquad \text{q. e. d.}$$

For $m = 0$ the result follows immediately from (II).

COROLLARY 3. 1. *$H^{m+n}(D, M) = H^m(D, M_n)$ for $m > 0$, $n \geq 0$.*

COROLLARY 3. 2. *$H^{m+n}(D, M) = H^m(D, M'_n)$ for $m > 0$, $n \geq 0$.*

4. The cohomology group of a Kronecker product.

LEMMA 4. 1. *Suppose $p > 0$, $f \in C^{p+1}(A \times B, M)$ and that*

(i) *f is a coboundary on B, i. e. there exists $g \in C^p(B, M)$ such that $\{f - \delta g\}({}_1b_{p+1}) = 0$.*

(ii) $$i > p,\ S_i \in B \Longrightarrow \delta f({}_1S_{p+2}) = 0.$$

Then there exists $h \in C^p(A \times B, M)$ such that

$$i > p,\ S_i \in B \Longrightarrow \{f - \delta h\}({}_1S_{p+1}) = 0.$$

Proof. Adopting the notation

$[\delta f({}_1S_{p+2})]_k$: the sum of the first k terms of the expansion of $\delta f({}_1S_{p+2})$,

${}_l[\delta f({}_1S_{p+2})]$: the sum of the l-th term and its successors in the expansion of $\delta f({}_1S_{p+2})$,

a_{np}: the product $a_n a_{n+1}\cdots a_p$, $(n \leq p)$,

and defining the cochains

$$u({}_1S_p) = a_{1p}\cdot g({}_1b_p) - f(a_{1p}, {}_1b_p),$$
$$v({}_1S_p) = (-1)^p f({}_1S_p, 1),$$
$$f_r({}_1S_p) = f({}_1S_{r-1}, a_{rp}, {}_rb_p), \qquad (1 < r \leq p),$$

we derive the following relations:

$$(1).\quad \delta u(S_1, b_2) = S_1\cdot u(b_2) - u(S_1b_2) + u(S_1)\cdot b_2$$
$$= S_1\cdot g(b_2) - S_1\cdot f(1, b_2) - a_1\cdot g(b_1b_2) + f(a_1, b_1b_2) + a_1\cdot g(b_1)\cdot b_2 - f(a_1, b_1)\cdot b_2$$
$$= a_1\cdot\delta g(b_1, b_2) - [S_1\cdot f(1, b_2)] + f(a_1, b_1)\cdot b_2$$
$$= \delta f(a_1, b_1, b_2) + f(S_1, b_2) - [\delta f(S_1, 1, b_2) - f(S_1, 1)\cdot b_2]$$
$$= f(S_1, b_2) - f(S_1, 1)\cdot b_2. \qquad (p = 1)$$

$$(2)\quad \delta u({}_1S_p, b_{p+1}) = S_1\cdot u({}_2S_p, b_{p+1}) - u(S_1S_2, {}_3S_p, b_{p+1}) + \cdots$$
$$+ (-1)^p u({}_1S_{p-1}, S_pb_{p+1}) - (-1)^p u({}_1S_p)\cdot b_{p+1}$$
$$= S_1a_{2p}\cdot g({}_2b_{p+1}) - a_{1p}\cdot g(b_1b_2, {}_3b_{p+1}) + \cdots$$
$$+ (-1)^p a_{1p}\cdot g({}_1b_{p-1}, b_pb_{p+1})$$
$$- (-1)^p a_{1p}\cdot g({}_1b_p)\cdot b_{p+1} - S_1\cdot f(a_{2p}, {}_2b_{p+1})$$
$$+ f(a_{1p}, b_1b_2, {}_3b_{p+1}) - \cdots - (-1)^p f(a_{1p}, {}_1b_{p-1}, b_pb_{p+1})$$
$$+ (-1)^p f(a_{1p}, {}_1b_p)\cdot b_{p+1}$$
$$= a_{1p}\cdot \delta g({}_1b_{p+1}) - S_1\cdot f(a_{2p}, {}_2b_{p+1}) + \delta f(a_{1p}, {}_1b_{p+1})$$
$$- a_{1p}\cdot f({}_1b_{p+1}) + f(a_{1p}b_1, {}_2b_{p+1})$$
$$= f(S_1a_{2p}, {}_2b_{p+1}) - S_1\cdot f(a_{2p}, {}_2b_{p+1}). \qquad (p>1)$$

$$(3).\quad \delta v({}_1S_p, b_{p+1}) = S_1\cdot v({}_2S_p, b_{p+1}) - v(S_1S_2, {}_3S_p, b_{p+1}) + \cdots$$
$$+ (-1)^p v({}_1S_{p-1}, S_pb_{p+1}) - (-1)^p v({}_1S_p)\cdot b_{p+1}$$
$$= (-1)^p[S_1\cdot f({}_2S_p, b_{p+1}, 1) - f(S_1S_2, {}_pS_3, b_{p+1}, 1) + \cdots$$
$$+ (-1)^p f({}_1S_{p-1}, S_pb_{p+1}, 1) - (-1)^p f({}_1S_p, 1)\cdot b_{p+1}]$$
$$= (-1)^p[\delta f({}_1S_p, b_{p+1}, 1) + (-1)^p f({}_1S_p, b_{p+1}) - (-1)^p f({}_1S_p, b_{p+1})\cdot 1$$
$$- (-1)^p f({}_1S_p, 1)\cdot b_{p+1}]$$
$$= f({}_1S_p, b_{p+1}) - f({}_1S_p, b_{p+1})\cdot 1 - f({}_1S_p, 1)\cdot b_{p+1}.$$

$$(4).\quad \delta f_p({}_1S_p, b_{p+1}) = S_1\cdot f_p({}_2S_p, b_{p+1}) - f_p(S_1S_2, {}_3S_p, b_{p+1}) + \cdots$$
$$- (-1)^p f_p({}_1S_{p-2}, S_{p-1}S_p, b_{p+1}) + (-1)^p f_p({}_1S_{p-1}, S_pb_{p+1})$$
$$- (-1)^p f_p({}_1S_p)\cdot b_{p+1}$$
$$= S_1\cdot f({}_2S_p, 1, b_{p+1}) - f(S_1S_2, {}_3S_p, 1, b_{p+1}) + \cdots$$
$$- (-1)^p f({}_1S_{p-2}, S_{p-1}S_p, 1, b_{p-1})$$
$$+ (-1)^p f({}_1S_{p-1}, a_p, b_pb_{p+1}) - (-1)^p f({}_1S_{p-1}, a_p, b_p)\cdot b_{p+1}$$
$$= \delta f({}_1S_p, 1, b_{p+1}) - (-1)^p f({}_1S_p, 1)\cdot b_{p+1}$$
$$+ (-1)^p f({}_1S_{p-1}, a_p, b_pb_{p+1}) - (-1)^p f({}_1S_{p-1}, a_p, b_p)\cdot b_{p+1}$$
$$= (-1)^p[f({}_1S_{p-1}, a_p, b_pb_{p+1}) - f({}_1S_p, 1)\cdot b_{p+1} - f({}_1S_{p-1}, a_p, b_p)\cdot b_{p+1}].$$

$$(5).\quad \delta f_r({}_1S_p, b_{p+1}) = S_1\cdot f_r(S_2, b_{p+1}) - f_r(S_1S_2, {}_3S_p, b_{p+1}) + \cdots$$
$$- (-1)^r f_r({}_1S_{r-2}, S_{r-1}S_r, {}_{r+1}S_p, b_{p+1})$$
$$+ (-1)^r f_r({}_1S_{r-1}, S_rS_{r+1}, {}_{p+1}S_{r+2}) + \cdots - (-1)^p f_r({}_1S_p)\cdot b_{p+1}$$
$$= S_1\cdot f\ ({}_2S_r, a_{(r+1)p}, {}_{r+1}b_{p+1}) - f(S_1S_2, {}_3S_r, a_{(r+1)p}, {}_{r+1}b_{p+1}) + \cdots$$
$$- (-1)^r f({}_1S_{r-2}, S_{r-1}S_r, a_{(r+1)p}, {}_pb_{r+1})$$
$$- (-1)^r f({}_1S_{r-1}, a_{rp}, b_rb_{r+1}, {}_{r+2}b_{p+1})$$
$$+ \cdots - (-1)^p f({}_1S_{r-1}, a_{rp}, {}_rb_p)\cdot b_{p+1}$$
$$= [\delta f({}_1S_r, a_{(r+1)p}, {}_{r+1}b_{p+1})]_r - {}_{r+2}[\delta f({}_1S_{r-1}, a_{rp}, {}_rb_{p+1})].$$

(6). Case 1. $p=2$. $\sum_{r=2}^{p}(-1)^r\delta f_r({}_1S_p, b_{p+1}) = \delta f_2({}_1S_2, b_3)$

$$
\begin{aligned}
&= f(S_1, a_2, b_2b_3) - f({}_1S_2, 1)\cdot b_3 - f(S_1, a_2, b_2)\cdot b_3\\
&= -\delta f(S_1, a_2, {}_2b_3) + S_1\cdot f(a_2, {}_2b_3) - f(S_1a_2, {}_2b_3)\\
&\qquad + f({}_1S_2, b_3) - f({}_1S_2, 1)\cdot b_3\\
&= S_1\cdot f(a_{2p}, {}_2b_{p+1}) - f(S_1a_{2p}, {}_2b_{p+1}) + f({}_1S_p, b_{p+1}) - f({}_1S_p, 1)\cdot b_{p+1}.
\end{aligned}
$$

(6). Case 2. $p>2$. $\sum_{r=2}^{p}(-1)^r\delta f_r({}_1S_p, b_{p+1})$

$$
\begin{aligned}
&= \{\delta f_2 - \delta f_3 + \cdots + (-1)^p\delta f_p\}({}_1S_p, b_{p+1})\\
&= \quad [\delta f({}_1S_2, a_{3p}, {}_3b_{p+1})]_2 \quad - \quad {}_4[\delta f(S_1, a_{2p}, {}_2b_{p+1})]\\
&\qquad - [\delta f({}_1S_3, a_{4p}, {}_4b_{p+1})]_3 \quad + \quad {}_5[\delta f({}_1S_2, a_{3p}, {}_3b_{p+1})]\\
&\qquad + [\delta f({}_1S_4, a_{5p}, {}_5b_{p+1})]_4 \quad - \quad {}_6[\delta f({}_1S_3, a_{4p}, {}_4b_{p+1})]\\
&\qquad\qquad \vdots \qquad\qquad\qquad\qquad \vdots\\
&+(-1)^p[\delta f({}_1S_{p-2}, a_{(p-1)p}, {}_{p-1}b_{p+1})]_{p-2} - (-1)^p{}_p[\delta f({}_1S_{p-3}, a_{(p-2)p}, {}_{p-2}b_{p+1})]\\
&-(-1)^p[\delta f({}_1S_{p-1}, a_p, {}_pb_{p+1})]_{p-1} \quad + (-1)^p{}_{p+1}[\delta f({}_1S_{p-2}, a_{(p-1)p}, {}_{p-1}b_{p+1})]\\
&+(-1)^p(-1)^p[f({}_1S_{p-1}, a_p, b_pb_{p+1}) - f({}_1S_p, 1)\cdot b_{p+1} - f({}_1S_{p-1}, a_p, b_p)\cdot b_{p+1}]\\
&= -{}_4[\delta f(S_1, a_{2p}, {}_2b_{p+1})]\\
&\qquad - f(S_1, S_2a_{3p}, {}_3b_{p+1}) \qquad\qquad + f({}_1S_2, S_3a_{4p}, {}_4b_{p+1})\\
&\qquad - f({}_1S_2, S_3a_{4p}, {}_4b_{p+1}) \qquad\qquad + f({}_1S_3, S_4a_{5p}, {}_5b_{p+1})\\
&\qquad\qquad \vdots \qquad\qquad\qquad\qquad \vdots\\
&\qquad - f({}_1S_{p-3}, S_{p-2}a_{(p-1)p}, {}_{p-1}b_{p+1}) + f({}_1S_{p-2}, S_{p-1}a_p, {}_pb_{p+1})\\
&\quad - (-1)^p[f({}_1S_{p-1}, a_p, {}_pb_{p+1})]_{p-1}\\
&\quad + f({}_1S_{p-1}, a_p, b_pb_{p+1}) - f({}_1S_p, 1)\cdot b_{p+1} - f({}_1S_{p-1}, a_p, b_p)\cdot b_{p+1}\\
&= S_1\cdot f(a_{2p}, {}_2b_{p+1}) - f(S_1a_{2p}, {}_2b_{p+1}) + f(S_1, S_2a_{3p}, {}_3b_{p+1})\\
&\quad - f(S_1, S_2a_{3p}, {}_3b_{p+1}) + f({}_1S_{p-2}, S_{p-1}a_p, {}_pb_{p+1})\\
&\quad - f({}_1S_{p-2}, S_{p-1}a_p, {}_pb_{p+1}) + f({}_1S_p, b_{p+1}) - f({}_1S_{p-1}, a_p, b_pb_{p+1})\\
&\qquad + f({}_1S_{p-1}, a_p, b_p)\cdot b_{p+1}\\
&\quad + f({}_1S_{p-1}, a_p, b_pb_{p+1}) - f({}_1S_p, 1)\cdot b_{p+1} - f({}_1S_{p-1}, a_p, b_p)\cdot b_{p+1}\\
&= S_1\cdot f(a_{2p}, {}_2b_{p+1}) - f(S_1a_{2p}, {}_2b_{p+1}) + f({}_1S_p, b_{p+1}) - f({}_1S_p, 1)\cdot b_{p+1}.
\end{aligned}
$$

It may now be verified by straightforward computation that the following definition of h will satisfy the requirements of the lemma:

For $p = 1$, $h(S_1) = \{2u - v\}(S_1) - u(S_1) \cdot 1$.

For $p > 1$, $h({}_1S_p) = \{2u - v + 2\sum_{r=2}^{p}(-1)^r f_r\}({}_1S_p)$

$$- \{u + \sum_{r=2}^{p}(-1)^r f_r\}({}_1S_p) \cdot 1.$$

This completes the proof of the lemma.

LEMMA 4. 2. *Suppose* $p \geq 0$, $B \varepsilon K_{p+1}$, $f \varepsilon C^{p+1}(A \times B, M)$, *and that* $i > p$, $S_i \varepsilon B \Longrightarrow \delta f({}_1S_{p+2}) = 0$. *Then there exists* $h \varepsilon C^p(A \times B, M)$ *such that* $i > p$, $S_i \varepsilon B \Longrightarrow \{f - \delta h\}({}_1S_{p+1}) = 0$.

Proof. The result is trivial for $p = 0$, and follows immediately from Lemma 4. 1 for $p > 0$.

LEMMA 4. 3. *Suppose* $p \geq 0$, $r \geq 1$, $B \varepsilon K_{p+1}$, $f \varepsilon C^{r+p+1}(A \times B, M)$ and *that*

(i) $$i > p,\ S_i \varepsilon B \Longrightarrow \delta f({}_1S_{r+p+2}) = 0,$$

(ii) $$i > p + 1,\ S_i \varepsilon B \Longrightarrow f({}_1S_{r+p+1}) = 0.$$

Then there exists $h \varepsilon C^{r+p}(A \times B, M)$ *such that*

$$i > p,\ S_i \varepsilon B \Longrightarrow \{f - \delta h\}({}_1S_{r+p+1}) = 0.$$

Proof. We apply section **3**, setting $D = A \times B$, $n = r$ and (in Prop. 3. 1) $m = p + 1$. We have, then, a cochain $\bar{f} \varepsilon C^{p+1}(A \times B, M_r)$ such that

(a) $$\{\delta\bar{f}({}_1S_{p+2})\}({}_{p+3}S_{r+p+2}) = \delta f({}_1S_{r+p+2}).$$

Now let N_r be the submodule of M_r consisting of all t such that

(b) $$i > p + 1,\ S_i \varepsilon B \Longrightarrow t({}_{p+2}S_{r+p+1}) = 0,$$

(c) $$i > p + 2,\ S_i \varepsilon B \Longrightarrow \{t \cdot S_{p+2}\}({}_{p+3}S_{r+p+2}) = 0.$$

We assert that $\bar{f} \varepsilon C^{p+1}(A \times B, N_r)$; for (b), in this case, follows immediately from (ii). To verify that f satisfies (a), we note first the following equality occurring in the proof of Prop. 3. 1:

$$\{\bar{f}({}_1S_{p+1}) \cdot S_{p+2}\}({}_{p+3}S_{r+p+2}) = (-1)^p[\delta f({}_1S_{r+p+2}) - S_1 \cdot f({}_2S_{r+p+2})$$
$$+ f(S_1S_2, {}_3S_{r+p+2}) - \cdots + (-1)^p f({}_1S_p, S_{p+1}S_{p+2}, {}_{p+3}S_{r+p+2}).$$

Now if $i > p+2$ and $S_i \varepsilon B$, we have the first term in the brackets zero by (i) and the remaining terms in the brackets zero by (ii), q. e. d.

Next we show that if we replace f and M in Lemma 4.1 by $\bar{f}$ and N_r respectively, then $\bar{f}$ satisfies the hypothesis of Lemma 4.1. First observe that (i) and (a) imply that $\bar{f}$ is a cocycle on B. But then, since $B \varepsilon K_{p+1}$, it follows that $\bar{f}$ is a coboundary on B. Next, note that the requirement $i > p$, $S_i \varepsilon B \Longrightarrow \delta\bar{f}({}_1S_{p+2}) = 0$ also follows from (i) and (a).

We may therefore conclude that for $p > 0$ there exists $\bar{h} \varepsilon C^p(A \times B, N_r)$ such that $i > p$, $S_i \varepsilon B \Longrightarrow \{\bar{f} - \delta\bar{h}\}({}_1S_{p+1}) = 0$; the same conclusion is trivial for $p = 0$.

Now we define $h \varepsilon C^{r+p}(A \times B, M)$ such that for $p > 0$: $h({}_1S_{r+p}) = \{\bar{h}({}_1S_p)\}({}_{p+1}S_{r+p})$; for $p = 0$: $h({}_1S_r) = \bar{h}({}_1S_r)$. To show that h satisfies the conditions of the lemma, we assert first that $i > p+1$, $S_i \varepsilon B \Longrightarrow \delta h({}_1S_{r+p+1}) = 0$; for we have $\delta h({}_1S_{r+p+1}) = \{\delta h({}_1S_{p+1})\}({}_{p+2}S_{r+p+1})$, and our assertion then follows from the definition of N_r.

Referring to (ii) we see that this proves the lemma for $i > p+1$, $S_i \varepsilon B$. There remains to prove only the case $i = p+1$, $S_i \varepsilon B$. But

$$\begin{aligned}\delta h({}_1S_p, b_{p+1}, {}_{p+2}S_{r+p+1}) &= \{\delta\bar{h}({}_1S_p, b_{p+1})\}({}_{p+2}S_{r+p+1}) \\ &= \{\bar{f}({}_1S_p, b_{p+1})\}({}_{p+2}S_{r+p+1}) = f({}_1S_p, b_{p+1}, {}_{p+2}S_{r+p+1}), \qquad \text{q. e. d.}\end{aligned}$$

LEMMA 4.4. *Suppose* $r \geq 0$, $p \geq 0$, $B \varepsilon K_{p+1}$, $f \varepsilon C^{r+p+1}(A \times B, M)$ *and*

$$i > p,\ S_i \varepsilon B \Longrightarrow \delta f({}_1S_{r+p+2}) = 0.$$

Then there exists $h \varepsilon C^{r+p}(A \times B, M)$ *such that*

$$i > p,\ S_i \varepsilon B \Longrightarrow \{f - \delta h\}({}_1S_{k+p+1}) = 0.$$

Proof. We use induction on r. For $r = 0$, the lemma coincides with Lemma 4.2. Suppose, then, that the lemma is true for $r = k \geq 0$, and that $r = k+1$. Replacing D in Prop. 3.2 by $A \times B$, and setting $n = 1$, $m = r + p = k + p + 1$, we have defined a cochain $\bar{f} \varepsilon C^{k+p+1}(A \times B, M'_1)$ such that $\{\delta\bar{f}({}_2S_{k+p+3})\}(S_1) = \delta f({}_1S_{k+p+3})$. But then $\bar{f}$ satisfies the hypothesis of the lemma for $r = k$, so that there exists by the inductive hypothesis a cochain $\bar{h}_1 \varepsilon C^{k+p}(A \times B, M'_1)$ such that

$$i > p,\ S_i \varepsilon B \Longrightarrow \{f - \delta\bar{h}_1\}({}_1S_{k+p+1}) = 0.$$

We now increase the dimension of $\bar{h}_1$ by defining $h_1 \varepsilon C^{k+p+1}(A \times B, M)$ such that $h_1({}_1S_{k+p+1}) = \{\bar{h}_1({}_2S_{k+p+1})\}(S_1)$ for $k + p > 0$, and $h_1(S_1) = \bar{h}_1(S_1)$ for $k + p = 0$. Then $\delta h_1({}_1S_{k+p+2}) = -\{\delta h_1({}_2S_{k+p+2})\}(S_1)$.

The cochain h_1 does not quite satisfy our requirements, since we now have only: $i > p+1, S_i \varepsilon B \Longrightarrow \{f - \delta h_1\}({}_1S_{k+p+2}) = 0$. However, this result means that $f - \delta h_1$ satisfies the hypotheses of Lemma 4.3 for $r = k+1$, with f replaced by $f - \delta h_1$. There exists therefore a cochain $h_2 \varepsilon C^{k+p+1}(A \times B, M)$ such that $i > p,\ S_i \varepsilon B \Longrightarrow \{f - \delta h_1) - \delta h_2\}({}_1S_{k+p+2}) = 0$.

Clearly, $h = h_1 + h_2$ satisfies the requirements of the lemma.

LEMMA 4.5. *Suppose* $r, p \geq 0$, $B \varepsilon K_{p+1}$, $\bar{f} \varepsilon C^{r+1}(A \times B, M'_p)$ *and that*

$$i > p,\ S_i \varepsilon B \Longrightarrow \delta \bar{f}({}_{p+1}S_{p+r+2}) = 0.$$

Then there exists $\bar{h} \varepsilon C^r(A \times B, M'_p)$ *such that*

$$i > p,\ S_i \varepsilon B \Longrightarrow \{\bar{f} - \delta \bar{h}\}({}_{p+1}S_{r+p+1}) = 0.$$

Proof. If $p = 0$ the lemma is a special case of Lemma 4.4. Suppose, then, that $p > 0$. We define f as in Prop. 3.2, replacing D by $A \times B$ and setting $m = r+1$, $n = p$. Then we have $\{\delta \bar{f}({}_{p+1}S_{p+r+2})\}({}_1S_p) = \delta f({}_1S_{r+p+2})$. Therefore $i > p,\ S_i \varepsilon B \Longrightarrow \delta f({}_1S_{r+p+2}) = 0$.

But then f satisfies the hypothesis of Lemma 4.4. Therefore there exists $h \varepsilon C^{r+p}(A \times B, M)$ such that $i > p,\ S_i \varepsilon B \Longrightarrow \{f - \delta h\}({}_1S_{r+p+1}) = 0$.

The cochain $\bar{h}$ defined as follows will now satisfy the requirements of the lemma: for $r > 0$, $\{\bar{h}({}_{p+1}S_{r+p})\}({}_1S_p) = (-1)^p h({}_1S_{r+p})$; for $r = 0$, $\bar{h}({}_1S_p) = (-1)^p h({}_1S_p)$.

THEOREM 4.1. $B \varepsilon K_{n+1} \Longrightarrow H^{m+n}(A \times B, M) \cong H^m[A, Z^0(B, M'_n)]$, $(m > 0, n \geq 0)$.

Proof. In Lemma 4.5, change the notation by replacing $\bar{f}$ by f, $\bar{h}$ by g, M'_p by P and $r+1$ by m. The result will be Lemma 3 of [5], (p. 574). As an immediate consequence of this lemma Hochschild proves ([5], Theorem 6, p. 575) the following result: $H^m[A, Z^0(B, M'_p)] \cong H^m(A \times B, M'_p)$, $(m > 0, p \geq 0)$. Theorem 4.1 now follows from Cor. 3.2.

5. The *K*- and *L*-classes of a Kronecker product.

THEOREM 5.1. $A \varepsilon K_m,\ B \varepsilon K_p \Longrightarrow A \times K_{m+p-1}$ $(m, p > 0)$.

Proof. Let M be any $A \times B$ module, $p = n+1$. Then the hypothesis of Theorem 4.1 is satisfied, so that $H^{m+n}(A \times B, M) \cong H^m[A, Z^0(B, M'_n)]$. But since $A \varepsilon K_m$ we have that $H^m[A, Z^0(B, M'_n)] = 0$. Therefore $H^{m+n}(A \times B, M) = 0$, i. e. $A \times B \varepsilon K_{m+n} = K_{m+p-1}$, q. e. d.

With reference to determining a lower bound for the *K*-class of $A \times B$,

only the two simplest cases are settled here. The result for the simplest case is given in Cor. 5.1. Lemma 5.1 introduces a general construction which seems to yield a result (Theorem 5.2) only for the next case.

PROPOSITION 5.1. *Let M be any B-module (B with or without an identity). We make P, the underlying vector space of $A \times M$, into an $A \times B$ module by the following definition:*

$$(a_1 \times b) \cdot (a_2 \times m) = a_1 a_2 \times b \cdot m, \quad (a_2 \times m) \cdot (a_1 \times b) = a_2 a_1 \times m \cdot b,$$

where $a_1, a_2 \varepsilon A$, $b \varepsilon B$ and $m \varepsilon M$ are elements of a pre-chosen basis for A, B, M respectively. This definition is then extended linearly to cover all of $A \times B$ and P.

Now let ϕ be the homomorphism of $C^k(B, M)$ into $C^k(A \times B, P)$ defined such that $\{\phi f\}({}_1S_k) = a_{1k} \times f({}_1b_k)$ for $k > 0$ and $\phi(m) = 1 \times m$ for $k = 0$.

Then ϕ induces an isomorphism of $H^k(B, M)$ into $H^k(A \times B, P)$, ($k \geq 0$).

Proof. First we show that $\phi\delta = \delta\phi$. For $k > 0$,

$$\begin{aligned}\{\delta(\phi f)\}({}_1S_{k+1}) &= S_1 \cdot \phi f({}_2S_{k+1}) - \phi f(S_1S_2, {}_3S_{k+1}) + \cdots \\ &\quad + (-1)^k \phi f({}_1S_{k-1}, S_kS_{k+1}) - (-1)^k \phi f({}_1S_k) \cdot S_{k+1} \\ &= a_{1k} \times [b_1 \cdot f({}_2b_{k+1}) - f(b_1b_2, {}_3b_{k+1}) + \cdots \\ &\quad + (-1)^k f({}_1b_{k-1}, b_kb_{k+1}) - (-1)^k f({}_1b_k) \cdot b_{k+1}] \\ &= a_{1k} \times \delta f({}_1b_{k+1}) = \{\phi(\delta f)\}({}_1S_{k+1}),\end{aligned}$$

proving our assertion for $k > 0$.

For $k = 0$: $\delta\{(\phi m)\}(S) = S \cdot \phi m - \phi m \cdot S = (a \times b) \cdot (1 \times m) - (1 \times m) = a \times \delta m(b) = \{\phi(\delta m)\}(S)$, proving our assertion for $k = 0$.

Now consider the mapping $f \to f \bmod B^k(A \times B, P)$. Since $\phi\delta = \delta\phi$, this mapping carries cocycles into cocycles and coboundaries into coboundaries. Thus all that remains to be proved is that the kernel of this mapping consists *only* of $B^k(B, M)$, in other words that if $\phi f = \delta g$, then there exists $h \varepsilon C^k(B, M)$ such that $f = \delta h$.

Let $1 = u_0, u_1 \cdots, u_r$ be a basis for A. We define a projection π of $A \times M$ into $1 \times M$ as follows. If $c \varepsilon A \times M$ and $c = \sum_{i=0}^{r} u_i \times m_i$, let $\pi c = 1 \times m_0$. Now, supposing $\phi f = \delta g$, let $1 \times h = \pi g$. Then

$$\phi f({}_1b_k) = 1 \times f({}_1b_k) = \delta g({}_1b_k) = 1 \times \delta h({}_1b_k).$$

Therefore $f = \delta h$, q. e. d.

COROLLARY 5.1.[2] $B \notin K_n \Longrightarrow A \times B \notin K_n$, ($B$ with or without identity, $n \geq 0$).

COROLLARY 5.2. $A \varepsilon L_0$, $B \varepsilon L_n \Longrightarrow A \times B \varepsilon L_n$, $(n \geq 0)$.

Proof of Corollary 5.2. $A \times B \notin K_n$ follows from Cor. 5.1; by Theorem 5.1 we have $A \times B \varepsilon K_{1+n+1-1} = K_{n+1}$, q. e. d.

We now define an algebraic "cup product" analogous to the cup product defined for groups by Eilenberg and MacLane. ([2], section 4).

Definition 5.1. Let A, B be algebras with or without identities, $f \varepsilon C^r(A, M)$, $g \varepsilon C^s(B, N)$, $(r, s > 0)$.[3]. Consider the vector space underlying $M \times N$ as an $A \times B$ module such that $(a \times b) \cdot (m \times n) = (a \cdot m) \times (b \cdot n)$, $(m \times n) \cdot (a \times b) = (m \cdot a) \times (n \cdot b)$.

Then the cup product $f \cup g \varepsilon C^{r+s}(A \times B, M \times N)$ is defined as follows:

$$\{f \cup g\}({}_1S_{r+s}) = [f({}_1a_r) \cdot a_{(r+1)(r+s)}] \times [b_{1r} \cdot g({}_{r+1}b_{r+s})].$$

It may be verified by straightforward computation that the cup product is associative, and we have

PROPOSITION 5.2. $\delta(f \cup g) = \delta f \cup g + (-1)^r f \cup \delta g$.

As a consequence of the associativity of the cup product we may extend Definition 5.1 inductively to any finite sequence of cochains; we note also that Proposition 5.2 implies that the cup product of cocycles is a cocycle and that the cup product of a cocycle and a coboundary (in either order) is a coboundary.

LEMMA 5.1. *Let $f \varepsilon Z^r(A, M)$. $g \varepsilon Z^s(B, N)$, $A \times B \varepsilon K_{r+s}$ $(r, s > 0)$, and suppose the identity elements of A, B act as identity operators on M, N respectively, and that f is "normalized" i. e. $1 \leq i \leq r$, $a_i = 1 \Longrightarrow f({}_1a_r) = 0$.*

Then given sequences $\alpha = {}_1a_r \varepsilon A$, $\beta = {}_1b_s \varepsilon B$, there exist cochains $f_\alpha \varepsilon C^{s-1}(B, M \times N)$ and $f_\beta \varepsilon C^{r-1}(A, M \times N)$ such that

$$f({}_1a_r) \times g({}_1b_s) = \delta f_\beta({}_1a_r) + \delta f_\alpha({}_1b_s).$$

Proof. Let $h = f \cup g$. By the remark preceding this lemma, h is a cocycle; hence, since $A \times B \varepsilon K_{r+s}$, h is a coboundary. Therefore there exists $\bar{h} \varepsilon C^{r+s-1}(A \times B, M \times N)$ such that $h = \delta \bar{h}$. Now consider

[2] For algebras A, B both with identities, this result is derived in Hochschild [5] as a consequence of Theorem 7, *ibid.*

[3] Although we shall not find it necessary to do so in the sequel, Def. 5.1 may be extended to include the cases $r = 0$, $s = 0$ also. In fact the cochain ϕf of Prop. 5.1 may be considered as a cup product in which $r = 0$, $s > 0$.

(I)

$$\begin{array}{llll}
h({}_1a_r, {}_1b_s) & = a_1 \cdot \bar{h}({}_2a_r, {}_1b_s) & - \cdots + (-1)^r \bar{h}({}_1a_{r-1}, a_r \times b_1, {}_2b_s) - \cdots \\
-h({}_1a_{r-1}, b_1, a_r, {}_2b_s) & = -a_1 \cdot \bar{h}({}_2a_{r-1}, b_1, a_r, {}_2b_s) & + \cdots - (-1)^r \bar{h}({}_1a_{r-1}, a_r \times b_1, {}_2b_s) - \cdots \\
\vdots & \vdots & \vdots
\end{array}$$

The arguments of h in the first column of (I) are to consist of the $(r+s)!/r!s!$ permutations of the sequence $({}_1a_r, {}_1b_s)$ which do not alter the order of the a's or the order of the b's. The signs at the extreme left of the rows of (I) are $+$ or $-$ according as the argument of h is an even or an odd permutation of $({}_1a_r, {}_1b_s)$.

Now we define $f_\beta = \bar{h}({}_1b_s)$ for $r = 1$ and $f_\alpha = (-1)^r \bar{h}({}_1a_r)$ for $s = 1$. For $r, s > 1$ we let $f_\beta({}_1a_{r-1}) = \bar{h}({}_1a_{r-1}, {}_1b_s) - \cdots$, where the terms on the right are of the form $\pm \bar{h}({}_1S_{r+s-1})$, with the sequences $({}_1S_{r+s-1})$ consisting of all permutations of $({}_1a_{r-1}, {}_1b_s)$ which do not alter the order of the a's or the order of the b's; the sign preceding $\bar{h}$ is to be $+$ or $-$ according as the argument of h is an even or an odd permutation of $({}_1a_{r-1}, {}_1b_s)$. Finally, let $f_\alpha = (-1)^r[\bar{h}({}_1a_r, {}_1b_{s-1}) - \cdots]$, with a similar convention as to the arguments of $\bar{h}$.

We assert that the sum of the right hand sides of the equation in (I) is $\delta f_\beta({}_1a_r) + \delta f_\alpha({}_1b_s)$. For firstly, if a summand on the right has an argument in which a term of the type $a_i \times b_j$ appear, then that summand appears precisely twice on the right and with opposite signs, so that all such summands disappear in adding the rows of (I); secondly, the remaining terms each occur precisely once and are clearly the terms of the expansions of $\delta f_\alpha({}_1b_s)$ and $\delta f_\beta({}_1a_s)$.

Finally, from the hypothesis of the lemma and the definition of $\bar{h}$, it follows that $h({}_1a_r, {}_1b_s) = f({}_1a_r) \times g({}_1b_s)$, while all the other terms on the left hand sides of the equations in (I) are zero. The sum of the left hand sides of the equations in (I) will then be $f({}_1a_r) \times g({}_1b_s)$, which completes the proof of the lemma.

LEMMA 5.2. *If $A \notin K_1$, then there exist an A-module M, $f \varepsilon Z^1(A, M)$ and $a_0 \varepsilon A$ such that the identity of A acts as an identity operator on M, $f(a_0) \neq 0$ and $m \varepsilon M \Longrightarrow a_0 \cdot m = m \cdot a_0$.*

Proof. Case 1. $R \neq 0$ is the radical of A and A/R is separable. Then $A = A/R + R$ as a supplementary sum, i. e. each element $a \varepsilon A$ may be uniquely expressed as $a = a' + r_a$, where $a' \varepsilon A/R$ and $r_a \varepsilon R$. Let π be the

projection of A onto R such that $\pi(a) = r_a$, let $M = R/R^2$, denote r/R^2 by $\bar{r}$, where $r \varepsilon R$, and define $a \cdot \bar{r} = \overline{ar}$, $\bar{r} \cdot a = \overline{ra}$, $f(a) = \overline{\pi a}$.

Then $1 \cdot \bar{r} = \bar{r} \cdot 1 = \bar{r}$; if we let a_0 be any element of $R - R^2$ then $f(a_0) = \overline{\pi a_0} = \bar{a}_0 \neq \bar{0}$; $\delta f(a'_1 + r_1, a'_2 + r_2) = \overline{a'_1 r_2 - (a'_1 r_2 + r_1 a'_2) + r_1 a'_2} = \bar{0}$, so that $f \varepsilon Z^1(A, M)$; finally, $a_0 \cdot \bar{r} = \bar{0} = \bar{r} \cdot a_0$. Thus all the requirements of the lemma are satisfied.

Case 2. $R = 0$.

In this case A is semisimple and inseparable. Let A_1 be a simple component of A of dimension m over its center C. It is shown in Hochschild [4], Lemma 4.1 that we may make the matrix algebra C_m into an A-module such that the identity of A acts as an identity operator on M, and such that there exists $a_0 \varepsilon C$ and $f \varepsilon Z^1(A, M)$ with $f(a_0) \neq 0$ and $a_0 \cdot m = m \cdot a_0$ for all $m \varepsilon M$. By allowing the other simple components of A to annihilate M and defining f to be zero on the other simple components of A, we make M into an A-module and extend f to A. The requirements of the lemma are then satisfied.

Case 3. $R \neq 0$ and A/R is inseparable.

Since A/R is semisimple, there exists by Case 2 an A/R module $\bar{M}$ and a cocycle $\bar{f} \varepsilon Z^1(A/R, \bar{M})$ satisfying the requirements of the lemma, with $A/R, \bar{M}, \bar{f}$ replacing A, M, F respectively. Letting $\bar{a}$ represent a/R, we may achieve the desired result by defining the vector space underlying M to coincide with that underlying $\bar{M}$, and for $m \varepsilon M$, $a \varepsilon A$ defining $a \cdot m = \bar{a} \cdot m$, $m \cdot a = m \cdot \bar{a}$, $f(a) = \bar{f}(\bar{a})$.

THEOREM 5.2. $A \notin K_1$, $B \notin K_n \Longrightarrow A \times B \notin K_{n+1}$, $(n > 0)$.

Proof. Let M, f, a_0 be as in Lemma 5.2, $g \varepsilon Z^n(B, N) - B^n(B, N)$, ${}_1b_n \varepsilon B$ and suppose $A \times B \varepsilon K_{n+1}$. Then by Lemma 5.1 there exist cochains $g_a \varepsilon C^n(B, M \times N)$, $f_\beta \varepsilon C^0(A, M \times N)$ such that $f(a) \times g({}_1b_n) = \delta f_\beta(a) + \delta g_a({}_1b_n)$.

We may assume (cf. Hochschild [5], section 1) that the identity element of B acts as an identity operator on N. Consequently Lemma 5.1 implies that $\delta f_\beta(a_0) = 0$. Now if we let $f(a_0) \times h({}_1b_{n-1})$ be a projection of $g_a({}_1b_{n-1})$ into $f(a_0) \times N$, we have $g = \delta h$, a contradiction which proves the theorem.

COROLLARY. $A \varepsilon L_1, B \varepsilon L_n \Longrightarrow A \times B \varepsilon L_{n+1}$, $(n \geq 0)$.

6. The existence of algebras in L_n, $(n \geq 0)$.

We now prove that none of the classes L_n is null. L_0 is not null, of

course, since $L_0 = K_1$ is the class of algebras separable over F. As for L_1, it is proved by Hochschild ([5], section 9) that $H' \varepsilon K_2$ where H' is an algebra with basis a, r such that a is a left identity and r a left annihilator. Since every separable algebra has an identity, it follows that $H' \not\varepsilon K_1$. Let H be the algebra formed by adjoining an identity 1 to H'. Then $H \varepsilon L_1$, since adjoining an identity to an algebra does not affect its K-class (cf. Hochschild [5], section 2).

PROPOSITION 6. 1.

$$H_1 = H_2 = \cdots = H_n = H \Longrightarrow H_1 \times H_2 \times \cdots \times H_n \not\varepsilon K_n.$$

Proof. Let R (with basis r) denote the radical of H. Let $R_1 = R_2 = \cdots = R_n = R$. We make $R_1 \times \cdots \times R_n$ into an $H_1 \times \cdots \times H_n$ module by defining $(h_1 \times \cdots \times h_n) \cdot (r_1 \times \cdots \times r_n) = h_1 r_1 \times \cdots \times h_n r_n$, $(r_1 \times \cdots \times r_n) \cdot (h_1 \times \cdots \times h_n) = r_1 h_1 \times \cdots \times r_n h_n$. Next define $f \varepsilon C^1(H, R)$ such that $f(1) = f(a) = 0$, $f(r) = r$. Then it is easily verified (either directly or from Lemma 5. 2, Case 1) that $f \varepsilon Z^1(H, R)$.

Now let g be the cup product $f_1 \cup \cdots \cup f_n$, where $f_1 = \cdots = f_n = f$. Then by the remark following Prop. 5. 2, g is a cocycle. If for $h_{ij} \varepsilon H$ we denote $h_{i1} \times \cdots \times h_{in}$ by U_i and $[h_{1i} \cdots h_{(i-1)i}][f(h_{ii})][h_{(i+1)i} \cdots h_{ni}]$ by V_i, then $g({}_1U_n) = V_1 \times \cdots \times V_n$.

In particular, if $h_{ii} = r$ and $h_{ij} = 1$ for $i \neq j$, then we denote U_i by W_i. Now, supposing that $g = \delta h$, consider the following set of equations (for convenience we let $W = W_1 W_2 = W_2 W_1$):

(II)

$$\begin{array}{rcl} g(W_1, W_2, {}_3W_n) & = & W_1 \cdot h(W_2, {}_3W_n) - h(W, {}_3W_n) + \cdots + (-1)^n h(W_1, W_2, {}_3W_{n-1}) \cdot W \\ -g(W_2, W_1, {}_3W_n) & = & -W_2 \cdot h(W_1, {}_3W_n) + h(W, {}_3W_n) - \cdots - (-1)^n h(W_2, W_1, {}_3W_{n-1}) \\ \vdots & & \vdots \qquad\qquad \vdots \qquad\qquad \vdots \end{array}$$

The arguments of g in the first column of (II) are to consist of the $n!$ permutations of the terms of the sequence $({}_1W_n)$. The signs at the extreme left of the rows of (II) are to be $+$ or $-$ according as the argument of g is an even or an odd permutation of $({}_1W_n)$.

We now note that each term in the first and last columns on the right hand side of the equations in (II) is zero, since each of these terms is equal to a Kronecker product with a factor in $R^2 = 0$. Secondly, note that each term in the columns between those just mentioned occurs precisely twice in

(II), but with opposite signs. The sum of the right hand terms of (II) is therefore zero. On the left hand side of the equations of (II), it is easily verified that all terms are zero except $g({}_1W_n) = r \times r \times \cdots \times r$.

We are therefore led by the assumption $g = \delta h$ to the contradiction $r \times r \times \cdots \times r = 0$. Thus g is a cocycle but not a coboundary; hence $H_1 \times \cdots \times H_n \not\in K_n$, q. e. d.

THEOREM 6.1. *L_n is not null,* $(n \geq 0)$.

Proof. We have already noted that $L_0 = K_1$ is not null. For $n > 0$, as an immediate consequence of Theorem 5.1 and Prop. 6.1, it follows that the algebra $H_1 \times \cdots \times H_n \in L_n$, where $H_1 = \cdots = H_n = H$.

UNIVERSITY OF MASSACHUSETTS.

BIBLIOGRAPHY

[1] S. Eilenberg, "Topological methods in abstract algebra," *Bulletin of the American Mathematical Society*, vol. 55 (1949), pp. 3-37.

[2] S. Eilenberg and S. MacLane, 'Cohomology theory in abstract groups. I," *Annals of Mathematics*, vol. 48 (1947), pp. 51-78.

[3] ———, "Cohomology theory in abstract groups. II," *ibid.*, vol. 48 (1947), pp. 326-341.

[4] G. Hochschild, "On the cohomology groups of an associative algebra," *ibid.*, vol. 46 (1945), pp. 58-67.

[5] ———, "On the cohomology theory for associative algebras," *ibid.*, vol. 47 (1946), pp. 568-579.

[6] ———, "Cohomology and representations of associative algebras," *Duke Mathematical Journal*, vol. 14 (1947), pp. 921-948.

ON RELATED PERIODIC MAPS.*

By E. E. Floyd.

1. **Introduction.** Consider a class of periodic maps defined on a topological space X. We are concerned with special cases of the following problem. Suppose the maps of the class are all related in some specified fashion. Are there, then, any implied relationships between the fixed point sets of the maps of the class?

A notable example of a problem of this sort has been solved recently by S. D. Liao [5]. If X is a finite dimensional compact Hausdorff space which has the homology groups of an n-sphere over the group I_p of integers mod p with p prime, and if T is periodic of period p on X, then, as P. A. Smith has proved ([8], p. 366), the fixed point set L has the homology groups of a r-sphere for some $-1 \leq r \leq n$. Liao settled a problem proposed by Smith by proving that if X also has finitely generated integral cohomology groups, then $n - r$ is even or odd according as T is orientation preserving or orientation reversing.

In section 1, we generalize Liao's result by proving that if X is a finite dimensional compact Hausdorff space with finitely generated integral cohomology groups, and if T is periodic of prime power period p^a on X, then the Lefschetz fixed point number of T is equal to the Euler characteristic of L (defined using I_p as coefficient group). We also extend a result of Smith ([9], p. 162) concerning the non-existence of certain types of periodic maps of arbitrarily large period on n-manifolds with negative Euler characteristic. The methods of this section depend heavily on recent results of Liao [5] and of the author [4] which in turn are based on the special homology groups of Smith [8].

In section 2, we consider a periodic map T of prime power period q^a and then consider the class of all periodic maps T_1 of the same period which are "sufficiently close" to T. Under these circumstances, we prove that the fixed point set L_1 of T_1 is close to L in the sense of Begle's metric [1] induced by the regular convergence introduced by Whyburn [11].

The author has read a pre-publication copy of Mr. Liao's paper [5], and wishes to thank Mr. Liao for that privilege.

* Received August 24, 1951; revised October 25, 1951.

2. The Lefschetz fixed point number of T. A periodic map on a space X generates a periodic linear isomorphism on the rational homology groups of X. We require later in the section an analysis of the latter. We dispose of this first, using a procedure similar to one used by Smith ([9], pp. 161-162) for a similar purpose.

Suppose V is a finite dimensional vector space over the rationals R. If W is a subspace of V, let dW denote the dimension of W. Let T be a linear transformation on V with $T^p =$ identity. There are associated with T the linear transformations $\sigma = 1 + T + \cdots + T^{p-1}$ and $\tau = 1 - T$. Clearly $\sigma\tau = \tau\sigma = 0$. We use the following preliminary remark (cf. [5], 4.11).

(2.1) *Image* $\sigma =$ *kernel* τ.

If m is a matrix presentation of T, then we call its characteristic equation $f(t)$ the characteristic equation of T. The characteristic roots of T are p-th roots of unity, for if $| m - \lambda I | = 0$, then $0 = | m^p - \lambda^p I | = (1 - \lambda^p)^{dV}$. Moreover, if no T^i, $0 < i < p$, has non-zero fixed points, then every characteristic root λ is a primitive p-th root of unity. For if $\lambda^l = 1$, then $| I - m^l | = | \lambda^l I - m^l | = 0$. Hence there exists $x \in V$, $x \neq 0$, with $T^l x = x$. But then $l = p$, so λ is a primitive p-th root.

Since $f(t)$ has rational coefficients and all its roots are p-th roots of unity, then $f(t) = f_{s_1}(t) \cdots f_{s_k}(t)$ where $f_{s_i}(t)$ is the cyclotomic equation of degree $\phi(s_i)$, and ϕ is Euler's ϕ-function, whose roots are the primitive s_i-th roots of unity. Moreover it may be seen that s_i divides p. In the following, we use $V(S)$ to represent the fixed point set of the linear transformation S.

(2.2) *Let T be a linear transformation on the finite-dimensional rational vector space V with $T^p =$ identity. Then*

(a) *if p is prime, there exists a non-negative integer k with $dV = dV(T) + k(p-1)$; moreover, trace $T = dV(T) - k$;*

(b) *if $p = q^a$ where q is prime and $a > 1$, then trace $T =$ trace $T \mid V(T^{q^{a-1}})$.*

Proof. To prove (a), decompose V into $V(T) \oplus V_1$, where $T(V_1) = V_1$ (cf. the proof of (2.1)). The characteristic equation of $T \mid V_1$ has as roots only primitive p-th roots of unity. Hence its characteristic equation is of the form $(f_p(t))^k$. Since the degree of $f_p(t)$ is $p-1$, $dV = dV(T) + k(p-1)$. The trace of $T \mid V_1$ is then $k(\alpha_1 + \cdots + \alpha_{p-1})$, where the α_i's are the primitive p-th roots of unity. Hence the trace of $T \mid V_1$ is $-k$. So (a) follows.

To prove (b), decompose V into $V(T^{q^{a-1}}) \oplus V_1$, where $T(V_1) = V_1$. Then the characteristic equation of $T \mid V_1$ is of the form $(f_p(t))^k$, and the trace of $T \mid V_1$ is $k(\alpha_1 + \cdots + \alpha_{\phi(p)})$, where the α_i's are the primitive p-th roots of unity. It may then be seen that the trace of $T \mid V_1$ is 0. So (b) follows.

Suppose now that X is a compact Hausdorff space, and let T be a map of X into X. Let $H_n(X;F)$ denote the Čech homology group of X over the field F, and T_{*n} the induced linear transformation on $H_n(X;F)$. Define $\chi(X;F) = \Sigma(-1)^i dH_i(X;F)$, in case the right hand side is defined and finite, and call $\chi(X;F)$ the Euler characteristic of X over F. Also define $\alpha(T;F) = \Sigma(-1)^i$ trace T_{*i}, in case $\chi(X;F)$ exists, and call $\alpha(T;F)$ the Lefschetz fixed point number of T over F ([6], p. 319).

We suppose now that X is a finite dimensional compact Hausdorff space with finitely generated integral Čech cohomology groups. Let T denote a periodic map on X of prime period p. Let L denote the fixed point set of T, and Y the orbit decomposition space of T. We have occasion to use the following recent results. Of these, (2. 3), (2. 4), and (2. 5) are due to Liao [5], and (2. 6) to the author [4].

(2. 3) (*Liao*). *Y has finitely generated cohomology groups.*

Liao ([5], Theorem 5. 5) has given a proof for this in case X has the groups of an n-sphere over I_p. The proof used the extra assumption only to insure that L has finitely generated groups over I_p. Since this is true in the general case ([4], Theorem 4. 2), the proof then holds.

(2. 4) (*Liao*). $\chi(X;I_p) = \chi(X;R)$, $\chi(Y;I_p) = \chi(Y;R)$ ([5], Theorem 2. 8).

(2. 5) *If $\eta: X \to Y$ denotes the orbit decomposition map, then η_* maps $[x \mid x \in H_n(X;R), T_* x = x]$ isomorphically onto $H_n(Y;R)$.*

This result is more or less implicit in the work of Liao (cf. [5], 4. 3, 4. 11, 4. 13). Because of its importance here, we outline, using the notation of [5; § 4], a direct argument. For each $b_{s\lambda} \in C_s(0(K_\lambda, T_\lambda);R)$, let $a_{s\lambda} \in C_s(K_\lambda, R)$ be such that $\eta_\lambda(a_{s\lambda}) = b_{s\lambda}$. Define $\xi_\lambda(b_{s\lambda}) = \sigma_\lambda a_{s\lambda}$. It may be verified that ξ_λ is uniquely defined, that $\partial\xi_\lambda = \xi_\lambda\partial$, and that $\pi_{\mu\lambda}\xi_\mu = \xi_\lambda\pi_{0\mu\lambda}$. Moreover, $\xi_\lambda\eta_\lambda = \sigma_\lambda$, and $\eta_\lambda\xi_\lambda(b_{s\lambda}) = pb_{s\lambda}$. Hence there is induced $\xi: H_s(0(X,T);R) \to H_s(X;R)$ with $\eta\xi(x) = px, x \in H_s(0(X,T);R)$, $\xi\eta(x) = \sigma(x), x \in H_s(X;R)$. Since $\eta\xi$ is an isomorphism onto, η maps image ξ isomorphically onto $H_s(0(X,T);R)$. Since η is onto and $\xi\eta = \sigma$, we have image $\xi =$ image σ. But by (2. 1) image $\sigma =$ kernel τ. The assertion follows.

(2.6) $\chi(X;I_p) + (p-1)\chi(L;I_p) = p\chi(Y;I_p)$. [4].

We are now in a position to prove the main theorem of this section.

(2.7) THEOREM. *Let X be a finite dimensional compact Hausdorff space with finitely generated integral Čech cohomology groups. Let T be a periodic map on X of period q^a, q prime. Let L be the fixed point set of T. Then $\alpha(T;R) = \chi(L;I_q)$.*

Proof. We prove the theorem first for $a = 1$. Consider $T_{*n}: H_n(X;R) \to H_n(X;R)$. According to (2.5), the fixed point set of T_{*n} is isomorphic to $H_n(Y;R)$. Hence by (2.5),

$$dH_n(X;R) = dH_n(Y;R) + [dH_n(Y;R) - \text{trace}\, T_{*n}](p-1)$$

so that $dH_n(X;R) + (p-1)\,\text{trace}\, T_{*n} = pdH_n(Y;R)$. Taking the alternating sum, we get $\chi(X;R) + (p-1)\alpha(T;R) = p\chi(Y;R)$. Using (2.4) and comparing with (2.6), we get $\alpha(T;R) = \chi(L;I_p)$.

Suppose $a > 1$ and suppose the theorem has been proven for $a - 1$. Consider $T_0 = T^{q^{a-1}}$. Let Y_1 denote the orbit space of the map T_0 on X, and $f: X \to Y_1$ the natural decomposition map. Define a map $S: Y_1 \to Y_1$ by $Sf = fT$. Then S is of period q^{a-1} on Y_1. Also, by (2.3), Y_1 has finitely generated integral cohomology groups. Hence, by the induction hypothesis, $\alpha(S;R) = \chi(L';I_q)$, where L' is the fixed point set of S.

We point out that L and L' are homeomorphic. Clearly, $f(L) \subset L'$ and f is 1-1 on L. We prove that $f(L) = L'$. Let $y \in L'$, where $y = f(x)$, $x \in X$. Then $f(x) = Sf(x) = fT(x)$ so $T(x) = T_0{}^k(x)$ for some k. But then $kq^{a-1} - 1$ is a period for x, so $kq^{a-1} - 1$ divides q^a. Hence $k = 0$, so that $Tx = x$, and $x \in L$. So $\chi(L';I_q) = \chi(L;I_q)$.

Finally, $\alpha(S;R) = \alpha(T;R)$. Let $F_n = [x; x \in H_n(X;R), T_{0*}x = x]$. Then, by (2.3), f_* maps F_n isomorphically onto $H_n(Y_1;R)$. Moreover, since $S_*f_* = f_*T_*$, we have trace $(S_{*n}; H_n(Y_1;R)) = \text{trace}(T_*, F_n)$. But by (2.2), $\text{trace}(T_*, F_n) = \text{trace}(T_*, H_n(X;R))$. It follows that $\alpha(S;R) = \alpha(T;R)$ and the theorem follows.

We now turn to some results concerned with properties of periodic maps of large period.

(2.8) (*Smith*). *Let V be a finite dimensional rational vector space. There exists a positive integer r associated with V so that if T is any linear transformation on V with $T^p =$ identity where $p > r$, then there exists $1 \leq j < p$ with $T^j =$ identity.*

Proof. We shall outline the proof ([9], pp. 161-162). Suppose $p = p_1^{a_1} p_2^{a_2} \cdots p_e^{a_e}$, where the p_i's are primes with $p_1 < p_2 < \cdots < p_e$. Define

$$\Phi(p) = \sum_1^e \phi(p_i^{a_i}) \text{ if } p_1 \neq 2 \text{ or } a_1 \neq 1;\ \Phi(p) = \sum_2^e \phi(p_i^{a_i}) \text{ otherwise.}$$

Then $\Phi(p) \to \infty$ as $p \to \infty$. We point out that if $\Phi(p) > dV$, then there exists $1 \leq j < r$ with $T^j =$ identity. For suppose this is not the case. Using the notation preceding (2.2), we have. $f(t) = f_{s_1}(t) \cdots f_{s_n}(t)$, where $s_i \mid p$. Now each $p_i^{a_i}$ divides some s_j. For if not, each s_j divides $p/p_i = q$, so that $T^q =$ identity. But if each $p_i^{a_i}$ divides some s_j, it may be checked that $dV = \Sigma\, \phi(s_i) \geq \Phi(p)$. Hence $\Phi(p) \leq dV$, and the assertion follows.

(2.9) *As a consequence of* (2.8), *let X be a compact Hausdorff space with each $H_n(X;R)$ of finite dimension and $= 0$ for all but a finite number of n's. There exists a positive integer r so that if T is any periodic map on X, then $T_*^j : H_n(X;R) \to H_n(X;R)$ is, for some $1 \leq j \leq r$, the identity for all n.*

We denote the least such r by $r(X)$.

(2.10) THEOREM. *Let X be a finite dimensional compact Hausdorff space with finitely generated integral cohomology groups. Let T be a periodic map on X of period $p > r(X)$. There exists $1 \leq i < p$ such that $p/i = q$ is a prime, and such that if L_i denotes the fixed point set of T^i, then $\chi(X;R) = \chi(L_i, I_q)$.*

Proof. There exists, by (2.9), $1 \leq j \leq r$ with $T_{*n}^j =$ identity for all n. Suppose $p = j \cdot k \cdot q$, where k and q are positive integers with q prime. Let $i = j \cdot k$. Then $T_{*n}^i =$ identity for all n. Hence by (2.7), $\alpha(T^i;R) = \chi(X;R) = \chi(L_i, I_q)$.

The following is an extension of a result of Smith [9; 162]. It also generalizes the well-known theorem [11] that the periodic maps on a compact 2-manifold with negative Euler characteristic have uniformly bounded periods. It does not, however, provide the upper bound known for that case.

(2.11) THEOREM. *Let X be a compact manifold with $\chi(X;R) < 0$. Suppose T is a periodic map on X of period p, and such that if $1 \leq j < p$, then the dimension of the fixed point set of T^j is ≤ 1. Then $p \leq r(X)$.*

Proof. Suppose $p > r(X)$. Let i be the number given by (2.10). Then $\chi(X;R) = \chi(L_i; I_q) < 0$. But $\dim L_i \leq 1$, so that by a result of

Smith ([10], p. 704), L_i is the union of a disjoint collection of points and simple closed curves. Hence $\chi(L_i; I_q) \geq 0$, which is a contradiction.

(2.12) *The above theorem is not true if the restriction on the dimension of the fixed point set of T^j is removed.*

As an example, let X be a 2-sphere, and let Y be a 2-manifold with $\chi(Y) < 0$. Then $\chi(X \times Y) = \chi(X)\chi(Y) < 0$. But since X admits transformations of arbitrary period, so does $X \times Y$.

3. Convergence properties. We begin section 3 by stating an important result due to Smith [7] which is the basis for the work of this section. The result is stated and proved in the proof of Theorems I, II in [7].

(3.1) (*Smith*). *Let X be a locally compact n-dimensional Hausdorff space, $n < \infty$, and let T be a periodic map on X of prime period p. Denote by L the fixed point set of T. Suppose $0 \neq A_0 \subset A_1 \subset \cdots \subset A_m$, $m = pn + p$, is a sequence of compact subsets of X, with $T(A_i) = A_i$, and with every Čech cycle in A_i over I_p bounding in A_{i+1}. Then $L \cap A_m \neq 0$ and every cycle in $L \cap A_0$ over I_p bounds in $L \cap A_m$.*

We use also the concept of regular convergence introduced by Whyburn [12]. We shall phrase the definition in terms of Čech theory instead of Vietoris theory; these are interchangeable, as follows from the full equivalence of the two theories ([6], p. 277). Let X be a locally compact metric space, and let G be an abelian group. Let $[A_i]$ be a sequence of closed subsets of X, with A_i converging to a closed subset A of X. If n is a non-negative integer, then A_i converges n-regularly to A over G if and only if given $x \in A$ and a compact neighborhood U of x in X, there exists a closed neighborhood V of x (in X) with $V \subset U$, and a positive integer I, so that every Čech cycle in $V \cap A_i$ over G of dimension $\leq n$ bounds in $U \cap A_i$ for $i > I$. It may be seen that X is lc^n (i. e., homologically locally connected over G in the dimensions from 0 to n), if and only if the sequence $X, X, \cdots$ converges n-regularly to X.

Let X and Y be metric spaces. Let A_i be a sequence of closed subsets of X which converges to a subset A of X. Let $f_i: A_i \to Y$, $f: A \to Y$ be continuous. We shall say that f_i *converges continuously* to f if and only if whenever $x_i \to x$, $x_i \in A_i$, then $f_i(x_i) \to f(x)$. This specializes, in case $A_i = A$, to the notion of continuous convergence introduced by Carathéodory ([2], p. 58).

(3.2) THEOREM. *Let X be a locally compact n-dimensional metric space, $n < \infty$. Suppose $[A_i]$ is a sequence of closed subsets of X converging n-regularly over I_p, p prime, to the subset A of X. Let T_i be a continuous periodic transformation of period p on A_i, such that $[T_i]$ converges continuously to the continuous function T on A. Then the fixed point set $[F_i]$ of T_i converges n-regularly over I_p to the fixed point set F of T.*

Proof. The reader may verify that if $x \in F$ and U is a neighborhood of x (in X), then there exists a neighborhood V of x and a positive integer I such that if $i > I$, then $\bigcup_j T_i^j(V \cap A_i) \subset U$.

Let $x \in F$ and let U be a compact neighborhood of x. There exists a sequence $U = U_{2m+1} \supset U_{2m} \supset \cdots \supset U_0$, $m = pn + p$, of compact neighborhoods of x (in X) and a positive integer I, such that $U_0 \cap A_i \neq 0$, for $i > I$, and (a) if $i > I$, then $\bigcup_j T_i^j(U_k \cap A_i) \subset U_{k+1}$ for $k = 0, \cdots, 2m$, and (b) for $i > I$ every cycle in $U_r \cap A_i$ over I_p bounds in $U_{r+1} \cap A_i$.

For each $0 \leq k \leq m$ and each $i > I$, define $V_{k,i} = \bigcup_j T_i^j(U_{2k} \cap A_i)$. Then $V_{k,i} \subset U_{2k+1} \cap A_i$, and $T_i(V_{k,i}) = V_{k,i}$. Moreover, since $V_{k+1,i} \supset U_{2k+2} \cap A_i$, every cycle in $V_{k,i}$ bounds in $V_{k+1,i}$. Hence we may apply (3.1) to the sequence $V_{0,i} \subset V_{1,i} \subset \cdots \subset V_{m,i}$, and the transformation T_i. It follows that $V_{m,i} \cap F_i \neq 0$, and every cycle in $V_{0,i} \cap F_i$ bounds in $V_{m,i} \cap F_i$. Hence, for $i > I$ every cycle in $U_0 \cap F_i$ bounds in $U \cap F_i$, and $U \cap F_i \neq 0$.

To finish the proof, the reader has only to note that if $x_{m_i} \in F_{m_i}$, and $x_{m_i} \to x$, then $x \in F$. This follows easily from continuous convergence.

(3.3) COROLLARY. *Let X be a locally compact n-dimensional metric space, $n < \infty$, which is lc^n over I_p, p prime. Let $[T_j]$ be a sequence of periodic maps on X of common period p^a, which converges continuously to the continuous map T. Then the fixed point set F_j of T_i converges n-regularly over I_p to the fixed point set F of T.*

Proof. The proof is a straight-forward combination of (3.2) together with a procedure used often by Smith for extending proofs from period p to period p^a ([8], p. 367).

(3.4) COROLLARY. *Let the hypotheses be those of* (3.3) *and suppose in addition that X is compact. Then there exists I such that for $i > I$, we have $H_j(F_i; I_p) \approx H_j(F; I_p)$ for all j. In particular, suppose X an n-sphere. Then there exists an integer r so that F_i, $i > I$, and F are all homological r-spheres over I_p.*

Proof. This follows from a theorem of Begle [1].

(3.5) COROLLARY. *Let X be an n-dimensional compact metric space, $n < \infty$, which is lc^n over I_p, p prime. Let T be a periodic map on X of period p^a with fixed point set L. There is an $\epsilon > 0$ such that if T_1 is periodic on X of period p^a, $\rho(T(x), T_1(x)) < \epsilon$ for all $x \varepsilon X$, and L_1 denotes the fixed point set of T_1, then $H_j(L; I_p) \approx H_j(L_1; I_p)$ for all j.*

UNIVERSITY OF VIRGINIA.

BIBLIOGRAPHY.

[1] E. G. Begle, "Regular convergence," *Duke Mathematical Journal*, vol. 11 (1944), pp. 441-450.

[2] C. Carathéodory, *Conformal representation*, Cambridge, 1932.

[3] E. E. Floyd, "Examples of fixed point sets of periodic maps," *Annals of Mathematics*, vol. 55 (1952), pp. 167-171.

[4] ———, "On periodic maps and the Euler characteristics of associated spaces," *Transactions of the American Mathematical Society*, vol. 72 (1952), pp. 138-147.

[5] S. D. Liao, "A theorem on periodic transformations of homology spheres," *Annals of Mathematics* (to appear).

[6] S. Lefschetz, *Algebraic topology*, New York, 1942.

[7] P. A. Smith, "Fixed point theorems for periodic maps," *American Journal of Mathematics*, vol. 63 (1941), pp. 1-8.

[8] ———, "Fixed points of periodic transformations," Appendix B in [6].

[9] ———, "Periodic and nearly periodic transformations," *Lectures in Topology*, Ann Arbor, 1941, pp. 159-190.

[10] ———, "Transformations of finite period II," *Annals of Mathematics*, vol. 39 (1938), pp. 127-164.

[11] F. Steiger, "Die maximalen Ordnungen periodischer topologischer Abbildungen geschlosser Flächen in sich," *Commentarii Mathematici Helvetici*, vol. 8 (1935), pp. 48-69.

[12] G. T. Whyburn, "On sequences and limiting sets," *Fundamenta Mathematicae*, vol. 25 (1935), pp. 408-426.

TOPOLOGY OF METRIC COMPLEXES.*

By C. H. Dowker.

The metric complexes (polyhedra) discussed in this paper are metric spaces with a cell decomposition and an affine structure for each cell. These complexes are subject to certain mild conditions (section 9, conditions a and b′) which, for example, ensure local connectedness. The complexes are not, however, required to be finite or countable. They may be curved and they need not be locally finite.

If a complex is star-finite, and if the closed cells are given their usual topology, then the topology of the whole complex is uniquely determined. However, if the complex is not star-finite, there is no longer a unique topolgy. J. H. C. Whitehead has chosen for his topological polyhedra ([11], pages 315-321) the finest topology consistent with the usual topology for the closed cells. This is a very convenient and useful topology, but with it all complexes except star-finiteones become non-metrizable spaces.

In connection with his studies of local connectedness, S. Lefschetz ([9], Chapter I) has chosen two particular ways of giving a complex a metric. If the complex is not star-finite, the topology induced by each of these metrics is necessarily less fine than the Whitehead topology. In general, the two different Lefschetz metrics induce different topologies.

In this paper, instead of choosing some particular metric, we take a somewhat axiomatic point of view and state conditions which should be satisfied by any metric complex. Our theorems are then shown to be consequences of these conditions. However, our method is not one of proof directly from the axioms. Instead, we use the method of comparing each metric complex with the corresponding Whitehead complex, that is, the same complex retopologized with the Whitehead topology.

In the first chapter, we discuss affine complexes. These are sets which have a cell decomposition and an affine structure for each cell, but they have no topology. These affine complexes have homology and cohomology groups, and the theorem on invariance under subdivision holds.

In the second chapter we add topology to the affine complex, and state conditions on the topology in order that the complex may be called a topological complex.

* Received August 10, 1950.

In the third chapter we discuss Whitehead complexes, that is, affine complexes with the Whitehead topology. Since the method of investigating metric complexes is that of comparison with the corresponding Whitehead complexes, we give a rather complete resumé of the known theorems on Whitehead complexes.

The fourth and main chapter contains the results on metric complexes. A metric complex is defined to be a topological complex whose topology is induced by a metric. Given a metric complex we construct (sections 10-13) a sequence of locally finite coverings of the complex by open sets, and using this sequence of coverings (sections 14-15) we construct a homotopy of the identity mapping of the complex. Then (section 16) by means of this homotopy, we prove that each metric complex has the same homotopy type as the corresponding Whitehead complex. It follows that any two isomorphic metric complexes have the same homotopy type. In section 17 we discuss the mappings of a space in the nerve of a covering when this nerve is a metric complex. In section 18 we show the topological invariance of the homology and cohomology groups of metric complexes.

I. Affine Complexes.

1. Definition and properties of affine complexes. By a convex cell E of a Euclidean space we mean an open bounded convex cell; its closure $\bar{E}$ is called a closed convex cell.

Let a set X be the union of a family $\{e_\alpha\}$ of mutually disjoint subsets e_α of K. Let each e_α be associated with a 1-1 transformation ϕ_α of some closed convex cell $\bar{E}_\alpha$ into X such that (i) ϕ_α maps the convex cell E_α onto e_α and (ii) if E' is any face of E, then $\phi_\alpha(E')$ is an e_β, and $\phi_\alpha^{-1}\phi_\beta$ is an affine (linear) transformation of $\bar{E}_\beta$ onto $\bar{E}'$. Then the set X, together with the decomposition $\{e_\alpha\}$ and the family $\{\phi_\alpha\}$ of transformations, is called an *affine complex.*

Each of the subsets e_α with the linear structure given by the transformation $\phi_\alpha \,|\, E_\alpha : E_\alpha \rightarrow e_\alpha$, is called a cell of the complex. The dimension of the cell e_α is defined to be the dimension of E_α. Each cell of dimension zero contains a single point which is called a vertex. If $e_\beta = \phi_\alpha(E')$, where E' is a face of E_α, e_β is called a face of e_α; we write $e_\beta \leq e_\alpha$. If E' is a proper face of E_α, e_β is called a proper face of e_α; we write $e_\beta < e_\alpha$. By a *finite* affine complex we mean one with only a finite number of cells. By a *star-finite* affine complex we mean an affine complex which, for each cell e_β, has only a finite number of cells e_α with $e_\beta \leq e_\alpha$.

If K is an affine complex, the cells e_α of K may be oriented by assigning orientations to the corresponding convex cells E_α. If e_β is an $(r-1)$-dimensional face of an r-cell e_α, the incidence number $[e_\alpha : e_\beta]$ is defined to be the incidence number $[E_\alpha : E']$, provided $\phi_\alpha^{-1}\phi_\beta : \bar{E}_\beta \to \bar{E}'$ is orentation preserving, and to be $-[E_\alpha : E']$ otherwise. With such a definition of orientation and incidence numbers, the affine complex becomes a closure finite oriented cell complex ([8], page 89). Thus, given a topological abelian group G and a non-negative integer p, we may define the p-dimensional cohomology group $H^p(K, G)$, and if G is discrete we may define the p-dimensional homology group $H_p(K, G)$.

The elements of the underlying set X of an affine complex K are called the points of K. A closed cell $\bar{e}_\alpha$ of K (the *closure* of e_α) is defined to be the image set $\phi_\alpha(\bar{E}_\alpha)$ together with the transformation ϕ_α. If x and y are points of $\bar{e}_\alpha$, the closed segment $[x, y]$ of $\bar{e}_\alpha$ is defined to be the image by ϕ_α of the closed segment of $\bar{E}_\alpha$ joining $\phi_\alpha^{-1}(x)$ to $\phi_\alpha^{-1}(y)$. If $0 \leq t \leq 1$, the point $tx + (1-t)y$ which divides the segment from x to y in $\bar{e}_\alpha$ in the ratio $t : 1-t$ is defined to be the image by ϕ_α of the point dividing the segment from $\phi_\alpha^{-1}(x)$ to $\phi_\alpha^{-1}(y)$ in the ratio $t : 1-t$. Similarly, a convex set in $\bar{e}_\alpha$ is defined to be the image by ϕ_α of a convex set of $\bar{E}_\alpha$. If A is a subset of $\bar{e}_\alpha$, the convex hull A^* of A in $\bar{e}_\alpha$ is the image by ϕ_α of the convex hull of $\phi_\alpha^{-1}A$ in $\bar{E}_\alpha$; thus A^* is the least convex set of $\bar{e}_\alpha$ containing A.

Note that $[x, y]$ may depend on $\bar{e}_\alpha$ as well as on x and y. However, if e_β is a face of e_α, and if x and y are points of $\bar{e}_\beta$, then the segment $[x, y]$ in $\bar{e}_\beta$ is the same as the segment $[x, y]$ in $\bar{e}_\alpha$, and the point $tx + (1-t)y$ in $\bar{e}_\beta$ is the same as the point $tx + (1-t)y$ in $\bar{e}_\beta$. In fact, since $\phi_\alpha^{-1}\phi_\beta : \bar{E}_\beta \to \bar{E}_\alpha$ is affine, the affine structure of $\bar{e}_\beta$ is that induced by the affine structure of $\bar{e}_\alpha$.

It is clear that the set $\bar{e}_\alpha$ is the union of all the faces of e_α; $\bar{e}_\alpha = \bigcup_{e_\beta \leq e_\alpha} e_\beta$. The star $\operatorname{St} e_\alpha$ of a cell e_α of K is defined to be the union of all cells e_β such that e_α is a face of e_β; $\operatorname{St} e_\alpha = \bigcup_{e_\alpha \leq e_\beta} e_\beta$. Note that $e_\alpha \subset \bar{e}_\alpha$, $e_\alpha \subset \operatorname{St} e_\alpha$, $e_\alpha = \bar{e}_\alpha \cap \operatorname{St} e_\alpha$. The following three statements are equivalent: (i) $e_\beta \leq e_\alpha$, (ii) $\bar{e}_\beta \subset \bar{e}_\alpha$, (iii) $\operatorname{St} e_\alpha \subset \operatorname{St} e_\beta$.

If x is a point of K, $e(x)$ is defined to be the unique cell e_α containing x, and $\bar{e}(x)$ is defined to be the closure of $e(x)$. The following five statements are equivalent: (i) $x \varepsilon \bar{e}(y)$, (ii) $\bar{e}(x) \subset \bar{e}(y)$, (iii) $e(x) \leq e(y)$, (iv) $\operatorname{St} e(y) \subset \operatorname{St} e(x)$, (v) $y \varepsilon \operatorname{St} e(x)$.

Two affine complexes K_1 and K_2 are called *isomorphic* if there is a 1-1 order preserving correspondence between the set of cells of K_1 and the set of

cells of K_2. The correspondence is called an isomorphism. It can be shown (See [1], page 127) that under an isomorphism, each cell corresponds to a cell of the same dimension.

Suppose there is given a subcollection of cells of an affine complex K such that if any cell e_α is in the subcollection, so is each face of e_α. Then the union L of the cells e_α of the subcollection is a set of points decomposed into cells e_α which are associated with transformations $\phi_\alpha : \bar{E}_\alpha \to L$. In fact L is an affine complex which is called a *subcomplex* of K. In particular, if e_α is a cell of K, the set $\bar{e}_\alpha$ with the obvious cell decomposition is a subcomplex of K. Also, the union of the closures of the cells contained in $\mathrm{St}\, e_\alpha$, with the obvious cell decomposition, forms a subcomplex which we call $\overline{\mathrm{St}}\, e_\alpha$.

If K and L are affine complexes, the product set $K \times L$ can be decomposed into cells $e_{\alpha\beta} = e_\alpha \times e_\beta$. For each $e_{\alpha\beta}$ let $\bar{E}_\alpha \times \bar{E}_\beta$ be the closed convex cell which is the cartesian product of $\bar{E}_\alpha$ and $\bar{E}_\beta$, and let $\phi_{\alpha\beta} : \bar{E}_\alpha \times \bar{E}_\beta \to K \times L$ be defined by $\phi_{\alpha\beta}(x, y) = (\phi_\alpha(x), \phi_\beta(y))$. It is easily verified that $K \times L$ thus becomes an affine complex. This complex is called the *product complex* of K and L.

An affine complex K is called *simplicial* if (i) for each e_α, E_α is a simplex, and (ii) each non-empty intersection $\bar{e}_\alpha \cap \bar{e}_\beta$ of two closed cells of K is a closed cell $\bar{e}_\gamma$. For a discussion of simplicial affine complexes see ([9], § 4).

2. Subdivision. A *subdivision* of an affine complex K is a 1-1 transformation $\mathrm{Sd} : K \to K'$ of K onto an affine complex K' such that (i) the image of each cell of K consists of the union of a finite number of cells, and (ii) the inverse transformation is linear on each closed cell of K'. We shall also say that the affine complex K' is a subdivision of K.

By condition (i), for each cell e_α of K' there is a unique cell e_β of K such that $e_\alpha \subset \mathrm{Sd}\, e_\beta$. Condition (ii) means that, if e_α and e_β have affine structures given by $\phi_\alpha : \bar{E}_\alpha \to \bar{e}_\alpha$ and $\phi_\beta : \bar{E}_\beta \to \bar{e}_\beta$, then $\phi_\beta^{-1} \mathrm{Sd}^{-1} \phi_\alpha : \bar{E}_\alpha \to \bar{E}_\beta$ is linear.

(2.1) *Isomorphic affine complexes have isomorphic simplicial subdivisions.*

Proof. The barycentric subdivision ([1], page 135) of an affine complex is a simplicial affine complex. Isomorphic affine complexes have isomorphic barycentric subdivisions.

(2.2) *If K and L are isomorphic simplicial affine complexes, there*

exists a map f of K onto L which maps each cell of K onto a corresponding cell of L, and which is linear on each closed cell.

Proof. The natural barycentric mapping ([9], p. 7; [1], p. 138, § 6) has the required properties.

Given a subdivision $\mathrm{Sd}: K \to K'$ of an affine complex K, let a chain transformation $\rho_p: C_p(K) \to C_p(K')$ be defined as follows: If e_β^p is an elementary p-chain of K, let $\rho_p e_\beta^p = \Sigma \epsilon_\alpha e_\alpha^p$, where the summation is over all α such that e_α^p is a p-cell of K', $e_\alpha^p \subset \mathrm{Sd}\, e_\beta^p$, and $\epsilon_\alpha = 1$ if $\phi_\beta^{-1}\mathrm{Sd}^{-1}\phi_\alpha$ is orientation preserving, $\epsilon_\alpha = -1$ if it is orientation reversing.

It is assumed known that, in the finite complex $\mathrm{Sd}\, \bar{e}_\beta$, $\rho\partial e_\beta^p = \partial\rho e_\beta^p$. Hence $\rho\partial = \partial\rho: C_p(K) \to C_{p-1}(K')$; thus ρ is a chain mapping ([6], p. 411; [8], p. 145). It is to be shown that ρ is a chain equivalence ([6], page 414).

If $e_\alpha \subset \mathrm{Sd}\, e_\beta$, let $\mathrm{T}\, e_\alpha$ be the finite subcomplex $\bar{e}_\beta$ of K. We define chain transformations $\tau_p: C_p(K') \to C_p(K)$ such that (i) $\tau_{p-1}\partial = \partial\tau_p: C_p(K') \to C_{p-1}(K)$, and (ii) $\tau_p e_\alpha^p$ is a chain of the subcomplex $\mathrm{T}\, e_\alpha^p$ of K. For each elementary 0-chain e_α^0, $\tau\, e_\alpha^0$ is chosen to be any elementary 0-chain in $\mathrm{T}\, e_\alpha^0$. If e_α^1 is an elementary 1-chain, $\tau\partial\, e_\alpha^1$ is a bounding 0-cycle in $\mathrm{T}\, e_\alpha^1$, and we chose as $\tau\, e_\alpha^1$ any 1-chain in $\mathrm{T}\, e_\alpha^1$ bounded by $\tau\partial\, e_\alpha^1$. If $p > 1$, assume that τ has been defined for dimensions less than p. Then $\tau\partial\, e_\alpha^p$ has been defined and is a chain in $\mathrm{T}\, e_\alpha^p$. Since $\partial\tau\partial\, e_\alpha^p = \tau\partial\partial\, e_\alpha^p = 0$, $\tau\partial\, e_\alpha^p$ is a cycle in $\mathrm{T}\, e_\alpha^p$; hence, since $\mathrm{T}\, e_\alpha^p$ is acyclic in dimension greater than zero, $\tau\partial\, e_\alpha^p$ is a bounding cycle in $\mathrm{T}\, e_\alpha^p$. Let $\tau\, e_\alpha^p$ be chosen as a chain of $\mathrm{T}\, e_\alpha^p$ whose boundary is $\tau\partial\, e_\alpha^p$.

In particular, if K and K' are simplicial, there is a simplicial map $\pi: K' \to \mathrm{K}$, called a projection, which maps each vertex v of K' into a vertex of $\mathrm{T}\, v$. Then π induces chain mappings $\pi_p: C\ (K') \to C_p(K)$ such that $\pi_p\, e_\alpha^p$ is a chain of $\mathrm{T}\, e_\alpha^p$. We may then take $\tau_p = \pi_p$.

It is to be shown that $\tau_p\rho_p: C_p(K) \to C_p(K)$ is the identity. This is clear for $p = 0$. Assume that $p > 0$ and that it is proved for dimensions less than p. Then $\tau\rho\, e^p$ is a chain of $\mathrm{TSd}\, e_\alpha^p = \bar{e}_\alpha^p$, and $\partial\tau\rho\, e_\alpha^p = \tau\rho\partial\, e_\alpha^p = \partial\, e_\alpha^p$ by the induction hypothesis. But the only chain of $\bar{e}_\alpha^p$ with boundary $\partial\, e_\alpha^p$ is the chain e_α^p. Hence $\tau\rho\, e_\alpha^p = e_\alpha^p$. It follows that $\tau\rho$ is the identity chain mapping.

We now show that the chain mapping $\rho\tau$ is chain homotopic to the identity. We define a homomorphism $D_{p+1}: C_p(K') \to C_{p+1}(K')$ so that $\partial Dc^p + D\partial c^p = c^p - \rho\tau c^p$, and so that De_α^p is a chain of $\mathrm{SdT}\, e_\alpha^p$. If e_α^0 is an elementary 0-chain, $e_\alpha^0 - \rho\tau e_\alpha^0 - D\partial e_\alpha^0 = e_\alpha^0 - \rho\tau e_\alpha^0$ is a bounding 0-chain in $\mathrm{SdT}\, e_\alpha^0$. Let $D_1 e_\alpha^0$ be chosen as a 1-chain in $\mathrm{SdT}\, e_\alpha^0$ whose boundary is $e_\alpha^0 - \rho\tau_\alpha^0$. Suppose that, for $p > 0$, D has been defined for chains of

dimension less than p. Then $D\partial e_a{}^p$ has been defined. Let $c_p = e_a{}^p - \rho\tau e_a{}^p - D\partial e_a{}^p$. Then c_p is a p-chain in SdT $e_a{}^p$, and by computation one finds that $\partial c_p = 0$. Thus c_p is a p-cycle in SdT $e_a{}^p$, a subcomplex which is acyclic in dimensions greater than zero. Hence c_p is a bounding cycle in SdT $e_a{}^p$. Let $D_{p+1}e_a{}^p$ be chosen as a $(p+1)$-chain in SdT $e_a{}^p$ bounded by c_p.

Thus $\tau\rho$ is the identity and $\rho\tau$ is chain homotopic to the identity. Therefore, ρ is a chain equivalence, and we have

(2.3) *If K' is a subdivision of an affine complex K, the homology group $H_p(K')$ is isomorphic with $H_p(K)$, and the cohomology group $H^p(K')$ is isomorphic with $H^p(K)$. If K and K' are simplicial, the homomorphisms $\pi_*: H_p(K') \to H_p(K)$ and $\pi^*: H^p(K) \to H^p(K')$ induced by the projection $\pi: K' \to K$ are isomorphisms onto.*

II. Topological complexes.

3. Definition of topological complexes. An affine complex K is called a topological complex if its underlying point set is a topological space, and if

(a) each $\phi_a: \bar{E}_a \to \bar{e}_a$ is a homeomorphism,

(b) for each neighborhood U of each point x of K, there is some neighborhood V of x such that, for each point y in V, $x \in \bar{e}(y)$, and the segment $[x, y]$ in $\bar{e}(y)$ is contained in U.

From condition (b) it follows immediately that as a space, K is locally connected. Also by condition (b), each point x has a neighborhood V such that, for each $y \in V$, $y \in \operatorname{St} e(x)$, hence such that $V \subset \operatorname{St} e(x)$. If e_a is any cell of K, and if $x \in \operatorname{St} e_a$, then $e_a \leq e(x)$, and $\operatorname{St} e(x) \subset \operatorname{St} e_a$. Thus each point x of $\operatorname{St} e_a$ has a neighborhood $V \subset \operatorname{St} e_a$; hence $\operatorname{St} e_a$ is open. Thus the star of any cell is an open set containing the cell. However, as the example below shows, one can have open stars and local connectedness with condition (b) not satisfied.

The complement in K of a closed cell $\bar{e}_a$ is the union of the stars of the cells in the complement; hence $K - \bar{e}_a$ is open, and $\bar{e}_a$ is a closed set in K. In fact, $\bar{e}_a$ is the topological closure in K of the subset e_a. By condition (a), $\bar{e}_a$ is compact.

Example. Let X be the subset of the cartesian plane with the following cell decomposition. The 0-cells of K are $A_0:(1,0)$ and $A_n:(2,1/n)$ for $n = 1, 2, \cdots$. The 1-cells of K are the segments A_nA_{n+1} and the broken lines A_0, $(-1/n, -1/n)$, $(-1/n, 1/n)$, A_n. The 2-cells of K are the regions

$A_0A_nA_{n+1}$ bounded by the 1-cells A_nA_{n+1}, A_0A_n and A_0A_{n+1}. Some suitable choice of the maps ϕ_α is to be made.

It may easily be verified that the space X of the example is locally connected, and that the star of each cell is an open set. However, condition (b) is not satisfied at the point A_0, and hence the complex is not a topological complex.

If K is any topological complex, an affine subcomplex is a topological subspace. Clearly conditions (a) and (b) hold also for the subcomplex. Thus a subcomplex of a topological complex is a topological complex.

If K_1 and K_2 are topological complexes, then $K_1 \times K_2$ is an affine complex and is also a topological space, the topological product of K_1 and K_2. Clearly $\phi_{\alpha\beta}: \bar{E}_\alpha \times \bar{E}_\beta \to \bar{e}_{\alpha\beta}$ $(= \bar{e}_\alpha \times \bar{e}_\beta)$ is a homeomorphism if $\phi_\alpha: \bar{E}_\alpha \to \bar{e}_\alpha$ and $\phi_\beta: \bar{E}_\beta \to \bar{e}_\beta$ are homeomorphisms. Thus condition (a) is satisfied. Any neighborhood of a point (x_1, x_2) of $K_1 \times K_2$ contains a neighborhood of the form $U_1 \times U_2$. Let neighborhoods V_1 of x_1 in K_1 and V_2 of x_2 in K_2 be chosen as in condition (b). Then $V_1 \times V_2$ is a neighborhood of (x_1, x_2) such that if $(y_1, y_2) \varepsilon V_1 \times V_2$, then $x_1 \varepsilon \bar{e}(y_1)$, $x_2 \varepsilon \bar{e}(y_2)$, and hence $(x_1, x_2) \varepsilon \bar{e}(y_1) \times \bar{e}(y_2) = \bar{e}(y_1, y_2)$, and the segment $[(x_1, x_2), (y_1, y_2)]$ in $\bar{e}(y_1, y_2)$ is contained in $[x_1, y_1] \times [x_2, y_2] \subset U_1 \times U_2$. Thus condition (b) is satisfied, and we have

(3.1) *the product of two topological complexes is a topological complex.*

III. Whitehead complexes.

4. The Whitehead topology. An arbitrary affine complex K can be made into a topological complex by giving it the finest[1] ([2], p. 9) topology consistent with condition (a). That is (cf. [11], p. 316), a set of K is called open if and only if the intersection with each closed cell $\bar{e}_\alpha$ is the image by ϕ_α of an open set of $\bar{E}_\alpha$. If U is any open set containing a given point x of K, let V be the set of points y of $\operatorname{St} e(x)$ such that the segment $[x, y]$ in $\bar{e}(y)$ is contained in U. It can be seen that the intersection of V with each closed cell $\bar{e}_\alpha$ is an open set of $\bar{e}_\alpha$; hence V is open. Clearly $x \varepsilon V$. Thus condition (b) is satisfied. This finest topology will be called the Whitehead topology, and the resulting topological complex will be called a Whitehead complex. It is known ([11], p. 320; [12], p. 225) that a Whitehead complex is a normal Hausdorff space.

[1] "Fine" in the sense of Bourbaki means "weak" in the sense of Whitehead or "strong" as used in functional analysis.

An equivalent description of the Whitehead topology is that a set of K is called closed if and only if its intersection with each closed cell $\bar{e}_a$ is the image by ϕ_a of a closed set of E.

It can be shown ([11], pp. 316-317) that if K is star-finite (in particular, finite), the Whitehead topology is the only one which satisfies conditions (a) and (b). Accordingly, given a star finite affine complex K, one may speak unambiguously of an open (or closed) set of K.

(4.1) *If L is a subcomplex of a Whitehead complex K, the subspace topology of L coincides with its Whitehead topology.*[2]

Proof. If A is a subset of L closed in the Whitehead topology of L, then $A \cap \bar{e}_a$ is closed in $\bar{e}_a$ for every $\bar{e}_a$ in L, hence also for every $\bar{e}_a$ in K. Therefore A is closed in K, and hence $A = A \cap L$ is closed in the subspace topology of L. Thus the Whitehead topology is not finer than the subspace topology, and the two topologies coincide.

(4.2) *A transformation f of a Whitehead complex K into a topological space Y is continuous if and only if $f|\bar{e}_a$ is continuous for each closed cell $\bar{e}_a$ of K.*

Proof.[3] If f is continuous, $f^{-1}(V)$ is open in K for each open set V of Y; hence $f^{-1}(V) \cap \bar{e}_a$ is open in $\bar{e}_a$, and $f|\bar{e}_a$ is continuous. Conversely, if $f|\bar{e}_a$ is continuous, $f^{-1}(V) \cap \bar{e}_a$ is open in $\bar{e}_a$, and hence $f^{-1}(V)$ is open in K; therefore f is continuous.

(4.3) *A transformation f of a Whitehead complex K into a topological space Y is continuous if and only if $f|L$ is continuous for each finite subcomplex L of K.*

Proof. If f is continuous, then $f|L$ is continuous for every subspace L, in particular for each finite subcomplex. If $f|L$ is continuous for each finite subcomplex L, then in particular $f|\bar{e}_a$ is continuous for each closed cell $\bar{e}_a$, and hence, by (4.2), f is continuous.

5. The product complex. If K is an affine complex, we denote by K_W the same complex with the Whitehead topology.

(5.1) *Let $K \times L$ be the product of two affine complexes. Then if either K or L is star-finite, $(K \times L)_W = K_W \times L_W$.*

[2] See Whitehead ([12], p. 224).

[3] See Whitehead ([11], p. 317; [12], p. 224).

Proof.[4] By (3.1), $K_W \times L_W$ is a topological complex. Hence the product topology is not finer than the Whitehead "finest" topology. It is then sufficient to show that if a set G is open in $(K \times L)_W$, it is open in $K_W \times L_W$; that is, for each point (x_0, y_0) of G there exist open sets U and V of K_W and L_W respectively, such that $x_0 \varepsilon U$, $y_0 \varepsilon V$ and $U \times V \subset G$.

Assume that it is L which is star-finite. For any cell $\bar{e}_\beta$ of L, $\bar{e}(x_0) \times \bar{e}_\beta = (\bar{e}(x_0) \times \bar{e}_\beta)_W$ is a finite subcomplex of $(K \times L)_W$, and $G_\beta = G \cap (\bar{e}(x_0) \times \bar{e}_\beta)_W$ is open in $\bar{e}(x_0) \times \bar{e}_\beta$. Hence $\{y \mid (x_0, y) \varepsilon G_\beta\}$ is open in $\bar{e}_\beta$. Let $H = \{y \mid (x_0, y) \varepsilon G\}$; then $H \cap \bar{e}_\beta = \{y \mid y \varepsilon \bar{e}_\beta, (x_0, y) \varepsilon G\} = \{y \mid (x_0, y) \varepsilon G_\beta\}$, which is open in $\bar{e}_\beta$. Hence H is open in L_W.

Clearly $y_0 \varepsilon H$. Since L_W is a normal space, we can choose an open set V of L_W so that $y_0 \varepsilon V$, $\bar{V} \subset H$, and $V \subset \operatorname{St} e(y_0)$. Let $M = \overline{\operatorname{St}}\, e(y_0)$; since L is star-finite, M is a finite subcomplex of L. Then $y_0 \varepsilon V \subset \bar{V} \subset H \cap M$. Let the subset U of K be defined by $U = \{x \mid x \times \bar{V} \subset G\}$. Since $\bar{V} \subset H$, $x_0 \varepsilon U$. Hence $(x_0, y_0) \varepsilon U \times V \subset G$. Thus there remains to be shown only that U is open in K_W.

If e_α is a cell of K, $\bar{e}_\alpha \times M = (\bar{e}_\alpha \times M)_W$ is a finite subcomplex of $(K \times L)_W$. Hence $G_\alpha = G \cap (\bar{e}_\alpha \times M)_W$ is an open set of $\bar{e}_\alpha \times M$. Then $U \cap \bar{e}_\alpha = \{x \mid x \varepsilon \bar{e}_\alpha, x \times \bar{V} \subset G\} = \{x \mid x \times \bar{V} \subset G_\alpha\}$, which is open in $\bar{e}_\alpha$, since G_α is open and $\bar{V}$ is compact. Hence U is open in the Whitehead complex K_W. This completes the proof of (5.1).

The condition of star-finiteness of one factor can not be dropped as the following example shows.

Example. Let K consist of a collection of closed 1-cells A_i of the power of the continuum, with a common vertex u_0. Let L be a countably infinite collection of closed 1-cells B_j, $j = 1, 2, \cdots$, with a common vertex v_0. Then $(K \times L)_W \neq K_W \times L_W$.

Proof. Let the closed 1-cells A_i have parameters x_i, $0 \leq x_i \leq 1$ so that, at the point u_0, $x_i = 0$ for all i. Let the closed 1-cells B_j have parameters y_j, $0 \leq y_j \leq 1$, and at v_0, let $y_j = 0$ for all j. Let the indices i be sequences of integers; $i = \{i_1, i_2, \cdots\}$. For each pair (i, j) of indices, let p_{ij} be the point $(1/i_j, 1/i_j)$ of $A_i \times B_j \subset K \times L$, and let P be the set of all such points p_{ij}. Then for each $A_i \times B_j$, $P \cap (A_i \times B_j)$ consists of one point p_{ij}, hence is closed in $A_i \times B_j$. Thus P is closed in $(K \times L)_W$.

A neighborhood U of u_0 in K_W is given by $x_i < a_i$, where the a_i are positive numbers. A neighborhood V of v_0 in L_W is given by $y_j < b_j$, $b_j > 0$.

[4] See Whitehead ([12], p. 227).

Let $U \times V$ be a product neighborhood of (u_0, v_0) in $K_W \times L_W$. Let $\bar{\imath} = \{\bar{\imath}_1, \bar{\imath}_2, \cdots\}$ be chosen so that for each j, $\bar{\imath}_j > j$ and $\bar{\imath}_j > (b_j)^{-1}$. Let $\bar{j}$ be chosen so $\bar{j} > (a_{\bar{\imath}})^{-1}$. Then $(\bar{\imath}_{\bar{j}})^{-1} < (\bar{j})^{-1} < a_{\bar{\imath}}$ and $(\bar{\imath}_{\bar{j}})^{-1} < b_{\bar{j}}$. Hence $p_{\bar{\imath}\bar{j}} \in U \times V$. Therefore every neighborhood of (u_0, v_0) in $K_W \times L_W$ contains a point of P. Hence, since $(u_0, v_0) \notin P$, P is not closed in the product topology. Thus the product topology is not so fine as the Whitehead topology.[5]

(5.2) *Let K be a Whitehead complex, let I be the closed interval $0 \leq t \leq 1$, and let Y be a space. Let h be a function from $K \times I$ to Y such that, for each closed cell $\bar{e}_a$ of K, $h \mid \bar{e}_a \times I$ is continuous. Then $h : K \times I \to Y$ is continuous.*[6]

Proof. We regard I as a complex consisting of one 1-cell and two 0-cells. Then by (5.1), since I is a finite (and hence star-finite) complex, $K \times I$ is a Whitehead complex. But for each closed cell $\bar{e}_\beta$ of $K \times I$, there is a closed cell $\bar{e}_a$ of K such that $\bar{e}_\beta \subset \bar{e}_a \times I$. Therefore, since $h \mid \bar{e}_a \times I$ is continuous, $h \mid \bar{e}_\beta$ is continuous. It follows from (4.2) that h is continuous.

6. Subdivisions of Whitehead complexes. One can show that any subdivision of a Whitehead complex is a Whitehead complex. More exactly, we have

(6.1) *If* $\mathrm{Sd} : K \to K'$ *is a subdivision of an affine complex K, and if both K and K' are given the Whitehead topology, then* Sd *is a homeomorphism.*

Proof. Sd^{-1} maps each closed cell of K' linearly, and hence continuously, into a closed cell of K. Hence Sd^{-1} is continuous. Sd maps each finite subcomplex L of K onto a finite subcomplex L' of K', and Sd/L is piecewise linear, hence continuous. Thus Sd is continuous, and therefore a homeomorphism.

(6.2) *Isomorphic Whitehead complexes are homeomorphic.*

Proof. Let K and L is isomorphic Whitehead complexes. If K and L are simplicial, the natural barycentric map of K on L is linear and hence continuous on each closed cell, and its inverse has the same property. Hence the natural barycentric map is a homeomorphism. If K and L are not simplicial, then by (2.1) they have isomorphic, and hence homeomorphic, simplicial subdivisions K' and L'. But by (6.1) K is homeomorphic to K', and L to L'. Hence K and L are homeomorphic.

---

[5] This answers a question of Whitehead, see [12, page 227, footnote 21].

[6] See Whitehead ([12], p. 228).

7. Coverings. A covering of a space X is a collection $\{U_a\}$ of open set whose union is X. A covering $\{U_a\}$ is called locally finite if every point of X has a neighborhood which meets only a finite number of sets U_a of the covering. A covering $\{V_\beta\}$ is called a refinement of $\{U_a\}$, if each V_β is contained in some U_a. If K is a topological complex, the star of each vertex v_a of K is an open set, and every point of K is in the star of some vertex; hence $\{\mathrm{St}\, v_a\}$ is a covering of K.

(7.1) *If K is a Whitehead complex and if $\{U_a\}$ is a covering of K. there exists a simplicial subdivision* $\mathrm{Sd}: K \to K'$ *such that the covering of K' by the stars of its vertices is a refinement of the covering* $\{\mathrm{Sd}\, U_a\}$.

For the proof see Whitehead ([11], Theorem 35).

Let X be a topological space, let $\mathfrak{U}$ be a covering of X, and let N $(= N(\mathfrak{U}))$ be the nerve of $\mathfrak{U}$ topologized as a Whitehead complex. Then N is a simplicial complex with a vertex u_a corresponding to each non-empty set of $\mathfrak{U}$, and a simplex $u_{a_0}, \cdots, u_{a_r}$ corresponding to each non-empty intersection $U_{a_0} \cap \cdots \cap U_{a_r}$. A mapping $\phi: X \to N$ is called canonical with respect to $\mathfrak{U}$ if for each $U_a \in \mathfrak{U}$, $\phi^{-1}\mathrm{St}\, u_a \subset U_a$.

(7.2) *If $\mathfrak{U}$ is a locally finite covering of a normal space X, there exists a canonical mapping of X into the Whitehead nerve N of $\mathfrak{U}$.*

Proof.[7] Let M be the nerve of $\mathfrak{U}$ metrized with the "natural" metric ([9], (4.12)). Then there exists ([4], 3(a); [3], Theorem 1.1) a canonical map θ of X into M. Now M and N are the same affine complex with two different topologies; let $\chi: M \to N$ be the identity map. Then χ is linear and hence continuous on each closed cell of M. Hence χ is continuous on each finite subcomplex of M. Hence by [3], Lemma 1.2, $\chi\theta: X \to N$ is continuous. Let $\phi = \chi\theta$; clearly ϕ is a canonical map of X into N.

8. Homology and cohomology groups. Let K be a Whitehead complex, and let $|K|$ be its underlying space. If K' is a subdivision of K, we identify the spaces $|K|$ and $|K'|$ by means of the homeomorphism $\mathrm{Sd}: |K| \to |K'|$. By the Čech cohomology groups of K we shall mean the Čech cohomology groups[8] of the space $|K|$.

(8.1) *The Čech cohomology groups of a Whitehead complex K are isomorphic with the corresponding combinatorial cohomology groups of K.*

---

[7] It may be seen that the proof of [3], Theorem 1.1 or the proof outlined for [4], proposition (a) does not depend on the special topology of the nerve; thus either of these gives a direct proof of (7.2).

[8] For definition and properties of Čech cohomology groups see [5].

Proof. Let K'_0 be a fixed simplicial subdivision of K. For each simplicial subdivision K'_α of K'_0, the nerve of the covering $\mathfrak{U}_\alpha$ of $|K|$ by the stars of the vertices of K'_α is a complex which can be identified with K'_α itself. By (7.1) these coverings $\mathfrak{U}_\alpha$ of $|K|$ by stars of vertices of simplicial subdivisions of K'_0 form a cofinal family of coverings. By (2.3), the projection map $\pi_{\alpha 0}$ of the nerve K'_α of $\mathfrak{U}_\alpha$ into the nerve K'_0 of $\mathfrak{U}_0$, induces an isomorphism $\pi^*_{\alpha 0}$ of $H^p(K'_0)$ onto $H^p(K'_\alpha)$. If K'_α and K'_β are two simplicial subdivisions of K'_0 such that $\mathfrak{U}_\beta$ is a refinement of $\mathfrak{U}_\alpha$, then ([5], p. 282) $\pi^*_{\beta 0} = \pi^*_{\beta\alpha}\pi^*_{\alpha 0}: H^p(K'_0) \rightarrow H^p(K'_\beta)$, and since $\pi^*_{\beta 0}$ and $\pi^*_{\alpha 0}$ are isomorphisms onto, $\pi^*_{\beta\alpha} = \pi^*_{\beta 0}\pi^*_{\alpha 0}{}^{-1}: H^p(K'_\alpha) \rightarrow H^p(K'_\beta)$ is also an isomorphism onto.

Since the coverings $\mathfrak{U}_\alpha$ form a cofinal family of coverings of $|K|$, the Čech cohomology group $H^p(|K|, G)$ of $|K|$ based on a discrete coefficient group G is the limit group of a direct spectrum, the groups of which are the cohomology groups $H^p(K'_\alpha, G)$ of the simplicial subdivisions of K_0 and the homomorphisms of which are the homomorphisms $\pi^*_{\beta\alpha}: H^p(K'_\alpha, G) \rightarrow H^p(K'_\beta, G)$. Since by (2.3) each $H^p(K'_\alpha, G)$ is isomorphic with $H^p(K, G)$, and since each $\pi^*_{\beta\alpha}$ is an isomorphism onto, it follows that the limit group $H^p(|K|, G)$ is also isomorphic with $H^p(K, G)$. Thus the Čech cohomology group $H^p(|K|, G)$ of K is isomorphic with the combinatorial cohomology group $H^p(K, G)$.

Another consequence of (7.1) is the following result.

(8.2) *The singular homology (cohomology) groups of a Whitehead complex K are isomorphic with the corresponding combinatorial homology (cohomology) groups of K.*

For the proof in case K is simplicial see ([7], pp. 399-400). If K is not simplicial, we replace it by a simplicial subdivision K'_0 where, by (6.1), $|K|$ and $|K'_0|$ are homeomorphic and, by (2.3), the combinatorial homology and cohomology groups of K are isomorphic with those of K'_0.

IV. Metric complexes.

9. Definition and properties of metric complexes. A metric complex is a topological complex whose underlying space is a metric space (a topological complex whose underlying space is metrizable will be called a metrizable complex). Replacing condition (b) of section **3** by the equivalent condition (b′), we can say that a metric complex is an affine complex K whose underlying set is a metric space subject to the conditions

(a) Each $\phi_\alpha: \bar{E}_\alpha \rightarrow \bar{e}_\alpha$ is a homomorphism,

(b′) For each point x of K and each positive number ϵ, there exists a positive number δ such that[9] if $\rho(x, y) < \delta$, then $x \in \bar{e}(y)$, and for each point z of the segment $[x, y]$ in $\bar{e}(y)$, $\rho(x, z) < \epsilon$.

Every subcomplex of a metric complex is a metric complex, for every subspace of a metric space is a metric space, and every subcomplex of a topological complex is a topological complex. The cartesian product of two metric complexes is a metric complex, for the product space of two metric spaces is a metric space, and the product of two topological complexes is a topological complex.

Let K be a metric complex. We define $\eta(x, \epsilon)$ as follows: Let $2\eta(x, \epsilon)$ be the least upper bound of the δ's of condition (b′) if this least upper bound exists and is less than ϵ; if the δ's are unbounded, or if their least upper bound is not less than ϵ, let $2\eta(x, \epsilon) = \epsilon$. Thus $\eta(x, \epsilon)$ is defined for each $x \in K$ and each $\epsilon > 0$, and $0 < \eta(x, \epsilon) \leq \epsilon/2$. Clearly if $\epsilon_1 < \epsilon_2$, then $\eta(x, \epsilon_1) \leq \eta(x, \epsilon_2)$. If $y \in K$, and $\rho(x, y) < 2\eta(x, \epsilon)$, then the segment $[x, y]$ exists in $\bar{e}(y)$, and for each $z \in [x, y]$, $\rho(x, z) < \epsilon$.

We define $\eta_r(x, \epsilon)$ inductively as follows, for $r = 0, 1, 2, \cdots$. Let $\eta_0(x, \epsilon) = \epsilon$, and for $r \geq 1$, let $\eta_r(x, \epsilon) = \eta(x, \eta_{r-1}(x, \epsilon))$. One sees immediately that $\eta_1(x, \epsilon) = \eta(x, \epsilon)$, and $\eta_2(x, \epsilon) = \eta(x, \eta(x, \epsilon))$. It also follows from the definition that for $r > 0$, $\eta_r(x, \epsilon) \leq \frac{1}{2}\eta_{r-1}(x, \epsilon)$, and for $r \geq 2$, $\eta_r(x, \epsilon) \leq \frac{1}{2}\eta(x, \epsilon)$.

(9.1) *For any metric complex, if* $r \geq 1$, *then* $\eta_r(x, \epsilon) = \eta_{r-1}(x, \eta(x, \epsilon))$.

Proof. This is clear for $r = 1$. We proceed by induction. Let $r \geq 2$, and assume $\eta_{r-1}(x, \epsilon) = \eta_{r-2}(x, \eta(x, \epsilon))$. Then $\eta_r(x, \epsilon) = \eta(x, \eta_{r-1}(x, \epsilon)) = \eta(x, \eta_{r-2}(x, \eta(x, \epsilon))) = \eta_{r-1}(x, \eta(x, \epsilon))$.

We write $S(x, \epsilon)$, or $S_0(x, \epsilon)$, for the set of points y of K such that $\rho(x, y) < \epsilon$, and $S_r(x, \epsilon)$ for the neighborhood $S(x, \eta_r(x, \epsilon))$. We write $\{p\}$ for the set consisting of one element p.

(9.2) *Let A be a set of r points* $(r \geq 1)$ *of a closed cell $\bar{e}_a$ of a metric complex K, let ϵ be a positive number, and let $\bar{x}$ be a point of A such that $\eta(\bar{x}, \epsilon) \geq \eta(x, \epsilon)$ for all $x \in A$. Then if there is some point y of $\bar{e}_a$ such that $y \in S_r(x, \epsilon)$ for all $x \in A$, the convex hull of $\{y\} \cup A$ in $\bar{e}_a$ is contained in $S(\bar{x}, \epsilon)$.*

Proof. First let $r = 1$. Then $A = \{\bar{x}\}$, and the convex hull of $\{y\} \cup A$ in $\bar{e}_a$ is the segment $[\bar{x}, y]$ in $\bar{e}_a$. Since $y \in S_1(\bar{x}, \epsilon) = S(\bar{x}, \eta(\bar{x}, \epsilon))$, $\rho(\bar{x}, y) < \eta(\bar{x}, \epsilon)$, and hence if $z \in [\bar{x}, y]$, $\rho(\bar{x}, z) < \epsilon$. Thus $[\bar{x}, y] \subset S(\bar{x}, \epsilon)$.

---

[9] We write $\rho(x, y)$ for the distance between the points x and y.

We proceed by induction. Let $r \geq 2$, and assume the result has been proved for $r-1$. Let $B = A - \{\bar{x}\}$, let $B_1 = \{y\} \cup B$, and let $A_1 = \{y\} \cup A$. Then for any point z of the convex hull A_1^* of A_1 in $\bar{e}_\alpha$, $z \in [\bar{x}, y_1]$, where y_1 is a point in the convex hull B_1^* of B_1 in $\bar{e}_\alpha$. By the induction hypothesis, since $y \in S_{r-1}(x, \eta(x, \epsilon)) \subset S_{r-1}(x, \eta(\bar{x}, \epsilon))$ for each $x \in B$, $B_1^* \subset S(\tilde{x}, \eta(\bar{x}, \epsilon))$ for some $\tilde{x} \in B$. Hence $\rho(\tilde{x}, y_1) < \eta(\bar{x}, \epsilon)$, and $\rho(\bar{x}, y_1) \leq \rho(\bar{x}, y) + \rho(y, \tilde{x}) + \rho(\tilde{x}, y_1) < \eta_r(\bar{x}, \epsilon) + \eta_r(\tilde{x}, \epsilon) + \eta(\bar{x}, \epsilon) < \frac{1}{2}\eta(\bar{x}, \epsilon) + \frac{1}{2}\eta(\tilde{x}, \epsilon) + \eta(\bar{x}, \epsilon) \leq 2\eta(\bar{x}, \epsilon)$. Therefore $\rho(x, z) < \epsilon$, and $A_1^* \subset S(\bar{x}, \epsilon)$.

(9.3) *For $r \geq 1$ let A be a set of r points, not necessarily distinct, of a closed cell $\bar{e}_\alpha$ of a metric complex K, and let ϵ be a positive number. Then if C is a non-empty convex set of $\bar{e}_\alpha$ contained in $\bigcap_{x \in A} S_r(x, \epsilon)$, the convex hull of $C \cup A$ in $\bar{e}_\alpha$ has diameter less than 2ϵ.*

Proof. Note that $(C \cup A)^*$ is the union of the sets $(\{y\} \cup A)^*$ for all $y \in C$. By Lemma 2, each $(\{y\} \cup A)^* \subset S(\bar{x}, \epsilon)$, where $\bar{x}$ is a point of A for which $\eta(\bar{x}, \epsilon)$ is maximal. Hence $(C \cup A)^* \subset S(\bar{x}, \epsilon)$, and the diameter of $(C \cup A)^*$ is less than 2ϵ.

10. The conditions to be satisfied. Let K be a metric complex. For each positive integer n we choose a collection $\mathfrak{U}^n$ of open sets of K, and with each open set U_λ of $\mathfrak{U}^n$ we associate a point x_λ of K. (The indices λ of the sets of $\mathfrak{U}^n$ are assumed to be elements of some sufficiently large index set.) With each cell e_α of K and each positive integer n, we associate a subcollection $\mathfrak{U}_\alpha^n$ of $\mathfrak{U}^n$ and a positive real number ρ_α^n. The choice of $\mathfrak{U}^n$, x_λ, $\mathfrak{U}_\alpha^n$ and ρ_α^n will be subject to the following conditions:

1) $\mathfrak{U}^n = \bigcup_\alpha \mathfrak{U}_\alpha^n$
2) $\mathfrak{U}_\alpha^n$ is a finite collection of open sets whose union contains $\bar{e}_\alpha$.
3) If $U_\lambda \in \mathfrak{U}_\alpha^n$ then $x_\lambda \in \bar{e}_\alpha$ and, for any $y \in U_\lambda$ and $z \in K - \operatorname{St} e(x_\lambda)$, $\rho(x_\lambda, y) < \frac{1}{2}\rho(x_\lambda, z)$.
4) If $e_\beta < e_\alpha$, $\mathfrak{U}_\beta^n \subset \mathfrak{U}_\alpha^n$.
5) If, for $e_\beta \leq e_\alpha$, $U_\lambda \in \mathfrak{U}_\alpha^n$, and $x_\lambda \in U_\mu \in \mathfrak{U}_\beta^n$, then $U_\lambda \in \mathfrak{U}_\beta^n$.
6) Let $e_\alpha \leq e_\gamma$, and let C be a convex set in $\bar{e}_\gamma$. Let $U_{\lambda_0}, \cdots, U_{\lambda_p}$

be sets of $\mathfrak{U}_\alpha^n$, and let $U_{\mu_0}, \cdots, U_{\mu_q}$ be sets of $\mathfrak{U}_\alpha^{n+1}$ such that the intersection

$$C \cap U_{\lambda_0} \cap \cdots \cap U_{\lambda_p} \cap U_{\mu_0} \cap \cdots \cap U_{\mu_q}$$

is not empty. Then if C has diameter less than $2\rho_\alpha^n$, the convex hull of the union $C \cup \{x_{\lambda_0}, \cdots, x_{\mu_q}\}$ in $\bar{e}_\gamma$ has diameter less than $1/n$.

Note that condition 3) implies condition 3′).

3′) If $U_\lambda \in \mathfrak{U}_\alpha^n$, then $e(x_\lambda) \leq e_\alpha$, and $U_\lambda \subset \mathrm{St}\, e(x_\lambda)$.

Also, if condition 3) holds for each of two cells e_α and e_β, then condition 3″) holds.

3″) If $U_\lambda \in \mathfrak{U}_\alpha^n$ and $U_\mu \in \mathfrak{U}_\beta^m$, and if $U_\lambda \cap U_\mu \neq 0$, then either

$$e(x_\mu) > e(x_\lambda) \quad \text{or} \quad e(x_\mu) \leq e(x_\lambda).$$

For if neither of $e(x_\lambda)$ and $e(x_\mu)$ is a face of the other, then $x_\lambda \in e(x_\lambda) \subset K - \mathrm{St}\, e(x_\mu)$, and $x_\mu \in e(x_\mu) \subset K - \mathrm{St}\, e(x_\lambda)$. Let $x \in U_\lambda \cap U_\mu$. By 3) applied to e_α, $\rho(x_\lambda, x) < \frac{1}{2}\rho(x_\lambda, x_\mu)$, and by 3) applied to e_β, $\rho(x_\mu, x) < \frac{1}{2}\rho(x_\mu, x_\lambda)$. Therefore $\rho(x_\lambda, x_\mu) > \rho(x_\lambda, x) + \rho(x_\mu, x)$, which is absurd.

Also, if conditions 2) and 5) hold for e_α and all of its faces, the following condition 5′) also holds.

5′) If, for $e_\beta \leq e_\alpha$, $U_\lambda \in \mathfrak{U}_\alpha^n$ and $x_\lambda \in \bar{e}_\beta$, then $U_\lambda \in \mathfrak{U}_\beta^n$.

For by 2). $\mathfrak{U}_\beta^n$ covers $\bar{e}_\beta$, and hence for some $U_\mu \in \mathfrak{U}_\beta^n$, $x_\lambda \in U_\mu$. Thus by 5), $U_\lambda \in \mathfrak{U}_\beta^n$.

11. The construction. We shall first construct the collections $\mathfrak{U}_\alpha^n$ of open sets, and afterwards we shall use condition 1 to define $\mathfrak{U}^n$. The construction of $\mathfrak{U}_\alpha^n$ will be by induction on the dimension of the cell e_α.

First let e_α be a cell of dimension zero; that is, e_α consists of a single point v. If v is the only point of K, let $\mathfrak{U}_\alpha^n$ consist of the one set $U_\lambda = \{v\}$, let x_λ be the point v, and let $\rho_\alpha^n = 1$. If K has other points, and hence other vertices, let $d(v)$ be the distance from v to the complement of the open set $\mathrm{St}\, e_\alpha$, and let $\mathfrak{U}_\alpha^n$ consist of a single open set U_λ which is the spherical neighborhood of v with radius the smaller of $\frac{1}{2}d(v)$ and $\eta(v, 1/2n)$. Let x_λ be the point v, and let $2\rho_\alpha^n = \eta(v, 1/2n)$. In either case the proof that conditions 2) to 6) are satisfied is easy, and is left to the reader.

Now suppose that e_α has dimension $k > 0$, and suppose that $\mathfrak{U}_\beta^n$, x_λ, and ρ_β^n satisfying conditions 2) to 6) have been constructed for all cells e_β of dimenion less than k (in particular for all proper faces e_β of e_α), and for all n.

For each n, let A_α^n be the set of points of $\bar{e}_\alpha$ which, for no proper face e_β of e_α, are contained in a set of the covering $\mathfrak{U}_\beta^n$. Then A_α^n is a closed and hence compact subset of $\bar{e}_\alpha$, and $A_\alpha^n \subset e_\alpha$. Let θ_α^n be the least value of ρ_β^m for all $e_\beta < e_\alpha$ and all $m \leq n$. Since e_α has only a finite number of faces, and since a finite number of positive integers precede $n + 1$, the number θ_α^n exists and is positive. For each point x of A_α^n, let $d(x)$ be the distance from

x to the complement of $\mathrm{St}\, e_\alpha$, and let $N(x)$ be the spherical neighborhood of x with radius $r(x)$ equal to the smaller of $\frac{1}{2}d(x)$ and $(1/3)\eta_{2k+2}(x, \theta_\alpha^n)$. Since A_α^n is compact and of dimension $\leq k$, the covering of A^n by the neighborhoods $N(x)$ has as a refinement a finite covering $\mathfrak{B}_\alpha^n$ by open sets of K, such that no point of K is contained in more that $k+1$ sets of $\mathfrak{B}_\alpha^n$. For each set U_λ of $\mathfrak{B}_\alpha^n$, let x_λ be any one of the points of A_α^n for which $N(x_\lambda) \supset U_\lambda$. Let κ_α^n be the least of the finite number of distances $r(x)$, more explicitly $r^n(x)$, corresponding to the finite number of U_λ in $\mathfrak{B}_\alpha^n$; and let ρ_α^n be the smaller of the two positive numbers κ_α^n and κ_α^{n+1}. Let $\mathfrak{U}_\alpha^n$ consist of the sets of $\mathfrak{B}_\alpha^n$ together with all sets of $\mathfrak{U}_\beta^n$ for all proper faces e_β of e_α.

12. Verification of the conditions. We now verify that the $\mathfrak{U}_\alpha^n$, x_λ, and ρ_α^n so defined satisfy conditions 2) to 6).

2) The collection $\mathfrak{U}_\alpha^n$ is the union of the finite collection $\mathfrak{B}_\alpha^n$ and the finite number of finite collections $\mathfrak{U}_\beta^n$ for $e_\beta < e_\alpha$. Hence $\mathfrak{U}_\alpha^n$ is finite. For any $x \,\varepsilon\, \bar{e}_\alpha$, either $x \,\varepsilon\, U_\lambda \,\varepsilon\, \mathfrak{U}_\beta^n$ for some $e_\beta < e_\alpha$, or $x \,\varepsilon\, A_\alpha^n$ and is contained in a set of the covering $\mathfrak{B}_\alpha^n$. Hence in either case, x is contained in a set of $\mathfrak{U}_\alpha^n$. Thus $\bar{e}_\alpha$ is contained in the union of the collection $\mathfrak{U}_\alpha^n$ of open sets, and condition 2) is satisfied.

3) If $U_\lambda \,\varepsilon\, \mathfrak{B}_\alpha^n$, then $x_\lambda \,\varepsilon\, A_\alpha^n \subset e_\alpha \subset \bar{e}_\alpha$. Then, if $y \,\varepsilon\, U_\lambda \subset N(x_\lambda)$, $\rho(x_\lambda, y) < r(x_\lambda) \leq \frac{1}{2}d(x_\lambda)$, and if $z \,\varepsilon\, K - \mathrm{St}\, e(x_\lambda) = K - \mathrm{St}\, e_\alpha$, $\rho(x_\lambda, z) \geq d(x_\lambda)$; hence $\rho(x, y) < \frac{1}{2}\rho(x, z)$. If $U_\lambda \,\varepsilon\, \mathfrak{U}_\alpha^n$ but $\not\varepsilon\, \mathfrak{B}_\alpha^n$, then for some $e_\beta < e_\alpha$, $U_\lambda \,\varepsilon\, \mathfrak{U}_\beta^n$. Applying condition 3) to the lower dimensional cell e_β, $x_\lambda \,\varepsilon\, \bar{e}_\beta \subset \bar{e}_\alpha$, and for $y \,\varepsilon\, U_\lambda$ and $z \,\varepsilon\, K - \mathrm{St}\, e(x_\lambda)$, $\rho(x_\lambda, y) < \frac{1}{2}\rho(x_\lambda, z)$. Thus condition 3) is satisfied.

4) By the definition of $\mathfrak{U}_\alpha^n$ condition 4) is satisfied.

5) Condition 5) is trivial for $e_\beta = e_\alpha$. Assume $e_\beta < e_\alpha$, $U_\lambda \,\varepsilon\, \mathfrak{U}_\alpha^n$, and $x_\lambda \,\varepsilon\, U_\mu \,\varepsilon\, \mathfrak{U}_\beta^n$. Then $x_\lambda \not\varepsilon\, A_\alpha^n$, and hence $U_\lambda \not\varepsilon\, \mathfrak{B}_\alpha^n$. Thus for some $e_\gamma < e_\alpha$, $U_\lambda \,\varepsilon\, \mathfrak{U}_\gamma^n$. Let $e_\delta = e(x_\mu)$. Then, by condition 3′) applied to e_β, since $U_\mu \,\varepsilon\, \mathfrak{U}_\beta^n$, we have $e_\delta \leq e_\beta$, and $U_\mu \subset \mathrm{St}\, e_\delta$. By condition 5′) applied to e_δ and e_β, since $U_\mu \,\varepsilon\, \mathfrak{U}_\beta^n$, and $x_\mu \,\varepsilon\, \bar{e}_\delta$, we have $U_\mu \,\varepsilon\, \mathfrak{U}_\delta^n$. By 3′) applied to e_γ, since $U_\lambda \,\varepsilon\, \mathfrak{U}_\gamma^n$, we have $e(x_\lambda) \leq e_\gamma$. But $x_\lambda \,\varepsilon\, U_\mu \subset \mathrm{St}\, e_\delta$; hence $e_\delta \leq e(x_\lambda)$, and $e_\delta \leq e_\gamma$. By 5) applied to e_δ and e_γ, since $U_\lambda \,\varepsilon\, \mathfrak{U}_\gamma^n$ and $x_\lambda \,\varepsilon\, U_\mu \,\varepsilon\, \mathfrak{U}_\delta^n$, we have $U_\lambda \,\varepsilon\, \mathfrak{U}_\delta^n$. Since $e_\delta \leq e_\beta$, it follows from condition 4) that $\mathfrak{U}_\delta^n \subset \mathfrak{U}_\beta^n$, Hence $U_\lambda \,\varepsilon\, \mathfrak{U}_\beta^n$, and condition 5) is satisfied.

6) Let $e_\alpha \leq e_\gamma$, and let C be a convex set in $\bar{e}_\gamma$ with diameter $< 2\rho_\alpha^n$. Let $y \,\varepsilon\, C \cap U_{\lambda_0} \cap \cdots \cap U_{\lambda_p} \cap U_{\mu_0} \cap \cdots \cap U_{\mu_q}$, where each U_λ is in $\mathfrak{U}_\alpha^n$, and each U_μ is in $\mathfrak{U}_\alpha^{n+1}$. Let u_n be the set $(x_{\lambda_0}, \cdots, x_{\lambda_p})$, and let u_{n+1} be the set $(x_{\mu_0}, \cdots, x_{\mu_q})$.

Each point of K, and in particular y, is contained in at most $k+1$ sets of $\mathfrak{B}_a{}^n$. Hence at most $k+1$ of the sets $U_{\lambda_0}, \cdots, U_{\lambda_p}$ are in $\mathfrak{B}_a{}^n$, and at most $k+1$ of the points of u^n are in e_a. Similarly, at most $k+1$ of the points of u^{n+1} are in e_a. Let $v^n = u^n \cap e_a$, $v^{n+1} = u^{n+1} \cap e_a$; then $v^n \cup v^{n+1}$ has at most $2k+2$ points.

If $x_\lambda \,\varepsilon\, v^n$, then $y \,\varepsilon\, U_\lambda \subset N(x_\lambda)$, and hence $\rho(x_\lambda, y) < r^n(x_\lambda)$. Also diam $C < 2\rho_a{}^n \leq 2\kappa_a{}^n \leq 2r^n(x_\lambda)$. Hence, if $z \,\varepsilon\, C$,

$$\rho(x_\lambda, z) \leq \rho(x_\lambda, y) + \rho(y, z) < 3r^n(x_\lambda) \leq \eta_{2k+2}(x_\lambda, \theta_a{}^n).$$

If $x_\mu \,\varepsilon\, v^{n+1}$, then $y \,\varepsilon\, U_\mu \subset N(x_\mu)$, and hence $\rho(x_\mu, y) < r^{n+1}(x_\mu)$. Also diam $C < 2\rho_a{}^n \leq 2\kappa_a{}^{n+1} \leq 2r^{n+1}(x_\mu)$. Hence if $z \,\varepsilon\, C$,

$$\rho(x_\mu, z) < 3r^{n+1}(x_\mu) \leq \eta_{2k-2}(x_\mu, \theta_a{}^{n+1}).$$

By definition, $\theta_a{}^n$ is a decreasing function of n; hence $\theta_a{}^{n+1} \leq \theta_a{}^n$, and $\eta_{2k+2}(x_\mu, \theta_a{}^{n+1}) \leq \eta_{2k+2}(x_\mu, \theta_a{}^n)$. Thus for each $z \,\varepsilon\, C$, $\rho(x_\mu, z) < \eta_{2k+2}(x_\mu, \theta_a{}^n)$.

It follows that for each of the at most $2k+2$ points x of $v^n \cup v^{n+1}$, C is contained in the $\eta_{2k+2}(x, \theta_a{}^n)$ neighborhood of x. Hence by (8.3), the diameter of the convex hull of $C \cup v^n \cup v^{n+1}$ in $\bar{e}_\gamma$ is less than $2\theta_a{}^n$.

Let $w^n = u^n - v^n$, and let $w^{n+1} = u^{n+1} - v^{n+1}$. Then for each $x_\nu \,\varepsilon\, w^n \cup w^{n+1}$, $e(x_\nu) < e_a$. Let x_π be chosen in $w^n \cup w^{n+1}$ so that $e(x_\pi)$ has maximum dimension, and let $e_\beta = e(x_\pi)$. Then if $x_\nu \,\varepsilon\, w^n \cup w^{n+1}$, $U_\pi \cap U_\nu \neq 0$, and hence by 3″) applied to the faces $e(x_\pi)$ and $e(x_\nu)$ of e_a, either $e(x_\nu) > e(x_\pi)$, or $e(x_\nu) \leq e(x_\pi)$. Since $e(x_\pi)$ is of maximal dimension, $e(x_\nu) > e(x_\pi)$ is impossible; hence $e(x_\nu) \leq e(x_\pi)$, and $x_\nu \,\varepsilon\, \bar{e}(x_\pi) = \bar{e}_\beta$. Therefore $w^n \cup w^{n+1} \subset \bar{e}_\beta$.

Let $C' = (C \cup v^n \cup v^{n+1})^*$, the convex hull of $C \cup v^n \cup v^{n+1}$. Then, diam $C' < 2\theta_a{}^n \leq 2\rho_\beta{}^n$. Now $C' \subset \bar{e}_\gamma$, $e_\beta < e_\gamma$, and y is in C' and in each of the sets U_ν of $(U_{\lambda_0}, \cdots, U_{\mu_q})$ for which $x_\nu \,\varepsilon\, w^n \cup w^{n+1} \subset \bar{e}_\beta$. But by condition 5′), each set U_ν is in $\mathfrak{U}_\beta{}^n \cup \mathfrak{U}_\beta{}^{n+1}$. Thus we can apply condition 6) to e_β, and we find that $(C' \cup w^n \cup w^{n+1})^*$ has diameter less than $1/n$. But

$$\begin{aligned}(C' \cup w^n \cup w^{n+1})^* &= ((C \cup v^n \cup v^{n+1})^* \cup w^n \cup w^{n+1})^* \\ &= (C \cup v^n \cup v^{n+1} \cup w^n \cup w^{n+1})^* = (C \cup u^n \cup u^{n+1})^*.\end{aligned}$$

Therefore $(C \cup u^n \cup u^{n+1})^* = (C \cup \{x_{\lambda_0}, \cdots, x_{\mu_q}\})^*$ has diameter less than $1/n$, and condition 6) is satisfied.

Thus for each e_a of dimension k, we construct $\mathfrak{U}_a{}^n$, $\{x_\lambda\}$, and $\rho_a{}^n$ satisfying conditions 2) to 6). By induction on k, we have a family of $\mathfrak{U}_a{}^n$, x_λ, and $\rho_a{}^n$ satisfying conditions 2) to 6) for all e_a and all n. Let $\mathfrak{U}^n = \bigcup_a \mathfrak{U}_a{}^n$. Then all the conditions 1) to 6) are satisfied.

13. The families of open sets as coverings. We first verify that each family $\mathfrak{U}^n$ is a covering and in fact a locally finite covering of the metric complex K.

(13.1) *Each of the families $\mathfrak{U}^n$, $n=1,2,\cdots$, of open sets is a locally finite covering of K.*

Proof. Let $x \varepsilon K$, and let $e_a = e(x)$. Then by condition 2), x is contained in some set of $\mathfrak{U}_a{}^n \subset \mathfrak{U}^n$. Therefore $\mathfrak{U}^n$ is a covering of K. Let W_a be the union of the open sets of $\mathfrak{U}_a{}^n$; then W_a is an open set containing x. Let $a(x)$ be a positive number less than half the distance from x to the complement of the open set $W_a \cap St\, e_a$, and let $G(x)$ be the spherical neighborhood of x with radius $a(x)$.

Let $U_\lambda \varepsilon \mathfrak{U}^n - \mathfrak{U}_a{}^n$, and let $e_\gamma = e(x_\lambda)$. Then for some e_δ, $U_\lambda \varepsilon \mathfrak{U}_\delta{}^n$, and $x_\lambda \varepsilon \bar{e}_\delta$. Hence by 5′), $U_\lambda \varepsilon \mathfrak{U}_\gamma{}^n$. Since $U_\lambda \not\varepsilon \mathfrak{U}_a{}^n$, e_γ is not a face of e_a, and so $e_a = e(x)$ is not in the star of $e_\gamma = e(x_\lambda)$. Hence $x \varepsilon K - \mathrm{St}\, e(x_\lambda)$.

Suppose it possible that $G(x) \cap U_\lambda \neq 0$, and let $y \varepsilon G(x) \cap U_\lambda$. Then by 3), since $y \varepsilon U_\lambda \varepsilon \mathfrak{U}_\gamma{}^n$, and $x \varepsilon K - \mathrm{St}\, e(x_\lambda)$, $\rho(x_\lambda, y) < \frac{1}{2}\rho(x_\lambda, x)$. Also, since $y \varepsilon G(x)$, $\rho(x,y) < a(x)$. Hence $\rho(x_\lambda, x) \leq \rho(x_\lambda, y) + \rho(x,y) < \frac{1}{2}\rho(x_\lambda, x) + a(x)$. Therefore $\rho(x_\lambda, x) < 2a(x)$, and hence $x_\lambda \varepsilon W_a \cap \mathrm{St}\, e_a$. Thus for some $U_\mu \varepsilon \mathfrak{U}_a{}^n$, $x_\lambda \varepsilon U_\mu$, and $e_a \leq e(x_\lambda) = e_\gamma$. Hence by 5), $U_\lambda \varepsilon \mathfrak{U}_a{}^n$, which is absurd. Therefore $G(x) \cap U_\lambda = 0$. Thus $G(x)$ meets only a finite number of open sets of $\mathfrak{U}^n$, $\mathfrak{U}^n$ is locally finite.

(13.2) *For any point $x \varepsilon K$, let u be the set of all x_λ for which $x \varepsilon U_\lambda \varepsilon \mathfrak{U}^n$, and let u' be the set of all x_μ for which $x \varepsilon U_\mu \varepsilon \mathfrak{U}^{n+1}$. Then u and u' are contained in $\bar{e}(x)$, and the convex hull of $\{x\} \cup u \cup u'$ in $\bar{e}(x)$ has diameter less than $1/n$.*

Proof. Let $e_a = e(x)$. and let C be the set $\{x\}$ consisting of one point. Then C is a convex set in $\bar{e}_a$, and diam $C = 0 < 2\rho_a{}^n$. As in the proof of (13.1), there is a neighborhood $G(x)$ of x which meets no sets of $\mathfrak{U}^n - \mathfrak{U}_a{}^n$. Hence the sets $U_{\lambda_0}, \cdots, U_{\lambda_p}$ of $\mathfrak{U}^n$ which contain x are in $\mathfrak{U}_a{}^n$, and $u = \{x_{\lambda_0}, \cdots, x_{\lambda_p}\}$ is contained in $\bar{e}_a$. Similarly, the sets $U_{\mu_0}, \cdots, U_{\mu_q}$ of $\mathfrak{U}^{n+1}$ which contain x are in $\mathfrak{U}_a{}^{n+1}$, and hence $u' = \{x_{\mu_0}, \cdots, x_{\mu_q}\}$ is contained in $\bar{e}_a$. Then by condition 6), the convex hull of $C \cup u \cup u' = \{x\} \cup u \cup u'$ in $\bar{e}_a$ has diameter less than $1/n$.

14. The mappings of K into K. Corresponding to each of the coverings $\mathfrak{U}^n$ of K we define a mapping $\psi_n : N^n \to K$, where N^n is the Whitehead nerve

of $\mathfrak{U}^n$. For each vertex u_λ of the nerve N^n there is a corresponding open set $U_\lambda \in \mathfrak{U}^n$, and there is a point x_λ of K associated with U_λ; we define $\psi_n(u_\lambda) = x_\lambda$. For each simplex $\sigma = u_{\lambda_0} \cdots u_{\lambda_p}$ of N^n, there is some point $x \in K$ contained in the intersection $U_{\lambda_0} \cap \cdots \cap U_{\lambda_p}$. Among the cells of K which contain at least one of the points $x_{\lambda_0}, \cdots, x_{\lambda_p}$, let e_μ be one of maximum dimension. Then it follows from condition 3″) that e_μ is unique, $x_{\lambda_0}, \cdots, x_{\lambda_p}$ are all in $\bar{e}_\mu$, and by condition 3′), $e_\mu \leq e(x)$. We define $\psi_n \mid \sigma$ to be the linear map of σ into $\bar{e}_\mu$ determined by mapping each vertex u_{λ_i} of σ into the corresponding x_{λ_i}. Since $e_\mu \leq e(x)$, $\psi_n \mid \sigma$ is also a linear map of σ into $\bar{e}(x)$ for each $x \in U_{\lambda_0} \cap \cdots \cap U_{\lambda_p}$. If σ_1 is a face of σ, the point x is in each U_{λ_i} corresponding to a vertex u_{λ_i} of σ_1, and therefore $\psi_n \mid \sigma_1$ is also a linear map into $\bar{e}(x)$. Hence $\psi_n \mid \bar{\sigma}$ is a linear map into $\bar{e}(x)$, and $\psi_n \mid \bar{\sigma}$ is continuous. Since N^n has the Whitehead topology, $\psi_n : N^n \to K$ is therefore continuous.

Let K_W be the complex K retopologized with the Whitehead topology, and let $f : K_W \to K$ be the identity map; i. e. for each $x \in K_W$, let $f(x) = x \in K$. Since the topology of K_W is at least as fine as that of K, f is continuous. The map $\psi_n : N^n \to K$ can be factored into a map $\chi_n : N^n \to K_W$ followed by $f : K_W \to K$. Since χ_n is linear and hence continuous on each closed simplex $\bar{\sigma}$ of N^n, $\chi_n : N^n \to K_W$ is continuous. Let ϕ_n be a canonical map of K into the nerve N^n of $\mathfrak{U}^n$. Then ([3], p. 202) for each $x \in K$, $\phi_n(x) \in \bar{\sigma}(x)$, where $\sigma(x)$ is the simplex of N^n determined by x; the vertices of $\sigma(x)$ correspond to the open sets of $\mathfrak{U}^n$ containing x. If $\sigma(x) = u_{\lambda_0} \cdots u_{\lambda_p}$, then $x \in U_{\lambda_0} \cap \cdots \cap U_{\lambda_p}$, and hence ψ_n maps $\bar{\sigma}(x)$ into $\bar{e}(x)$. Hence $\psi_n\phi_n(x) \in \bar{e}(x)$. Let $g_n = \chi_n\phi_n : K \to K_W$; then (see diagram) $fg_n = f\chi_n\phi_n = \psi_n\phi_n : K \to K$. Thus, for each $x \in K$, $fg_n(x) \in \bar{e}(x)$.

$$\begin{tikzcd} K \arrow[r, "\phi_n"] \arrow[d, "g_n"'] & N^n \arrow[d, "\psi_n"] \arrow[dl, "\chi_n"'] \\ K_W \arrow[r, "f"] & K \end{tikzcd}$$

15. **The homotopy.** We now define a homotopy $h : K \times I \to K$. For $n = 1, 2, \cdots$, let $h(x, 1/n) = fg_n(x) = \psi_n\phi_n(x)$. For $1/(n+1) < t < 1/n$, let $h(x, t)$ be the point of $\bar{e}(x)$ which divides the segment from $h(x, 1/(n+1))$ to $h(x, 1/n)$ in the ratio of $t - 1/(n+1)$ to $1/n - t$. Let $h(x, 0) = x$. We must show that $h : K \times I \to K$ is continuous.

Let $x \varepsilon K$, and let $e_a = e(x)$. Let V be an open set containing x which meets no sets of $\mathfrak{U}^n - \mathfrak{U}_a{}^n$ and meets no sets of $\mathfrak{U}^{n+1} - \mathfrak{U}_a{}^{n+1}$; then $V \subset \mathrm{St}\, e_a$. Let I_n be the interval $1/(n+1) \leq t \leq 1/n$. Let $(y, t) \varepsilon V \times I_n$. Then since $y \varepsilon V \subset \mathrm{St}\, e_a$, $e_a \leq e(y)$. If $y \varepsilon U_\lambda \varepsilon \mathfrak{U}^n$, we have $U_\lambda \varepsilon \mathfrak{U}_a{}^n$, and hence $x_\lambda \varepsilon \bar{e}_a$. Thus for each vertex u_λ of $\sigma(y)$ in N^n, $\psi_n(u_\lambda) \varepsilon \bar{e}_a$; hence ψ_n maps $\bar{\sigma}(y)$ into $\bar{e}_a \subset \bar{e}(y)$. Therefore, since $\phi_n(y) \varepsilon \bar{\sigma}(y)$, $fg_n(y) = \psi_n\phi_n(y) \varepsilon \bar{e}_a$. Similarly $fg_{n+1}(y) \varepsilon \bar{e}_a$. Since $e_a \leq e(y)$, the segment joining $fg_{n+1}(y)$ to $fg_n(y)$ in $\bar{e}(y)$ is the segment joining them in $\bar{e}_a$, and the point $h(y, t)$ is the point dividing the segment from $fg_{n+1}(y)$ to $fg_n(y)$ in $\bar{e}_a$ in the ratio $t - 1/(n+1)$ to $1/n - t$. Thus h maps $V \times I_n$ continuously into $\bar{e}_a \subset K$. It follows that for each n, h maps $K \times I_n$ continuously, and from this it follows that h is continuously except possibly for $t = 0$.

We must show that h is also continuous at points $(x, 0)$ of $K \times I$. For any $x \varepsilon K$ and $\epsilon > 0$, let U be the $(\epsilon/2)$-neighborhood of x in K, and let W be the subset $\{t \mid 0 \leq t < 1/m\}$ of I, where m is some integer greater than $2/\epsilon$. Then $U \times W$ is a neighborhood of $(x, 0)$ in $K \times I$. Let (y, t) be any point of $U \times W$. If $t = 0$, $h(y, t) = y$, $h(x, 0) = x$, and $\rho(h(x, 0), h(y, t)) = \rho(x, y) < \epsilon/2 < \epsilon$. If, on the other hand, $t > 0$, then for some $n \geq m$, $1/(n+1) \leq t \leq 1/n$. If $U_{\lambda_0}, \cdots, U_{\lambda_p}$ are the sets of $\mathfrak{U}^n$ containing y, then $\phi_n(y) \varepsilon \bar{\sigma}(y)$, where $\sigma(y)$ is the simplex $u_{\lambda_0}, \cdots, u_{\lambda_p}$ of N^n. Since $\psi_n(u_{\lambda_i}) = x_{\lambda_i}$, and since $\psi_n \mid \bar{\sigma}(y)$ is a linear map into $\bar{e}(y)$, ψ_n maps $\bar{\sigma}(y)$ into the convex hull of $\{x_{\lambda_0}, \cdots, x_{\lambda_p}\}$ in $\bar{e}(y)$. Similarly if $U_{\mu_0}, \cdots, U_{\mu_q}$ are the sets of $\mathfrak{U}^{n+1}$ containing y, $\psi_{n+1}\phi_{n+1}(y) \varepsilon \{x_{\mu_0}, \cdots, x_{\mu_q}\}^*$ in $\bar{e}(y)$. Therefore, since $h(x, t)$ is on the segment from $\psi_{n+1}\phi_{n+1}(y)$ to $\psi_n\phi_n(y)$ in $\bar{e}(y)$, $h(x, t) \varepsilon \{x_{\lambda_0}, \cdots, x_{\mu_q}\}^*$ in $\bar{e}(y)$. By (12.2), diam $\{y, x_{\lambda_0}, \cdots, x_{\mu_q}\}^* < 1/n$; hence $\rho(y, h(x, t)) < 1/n \leq 1/m < \epsilon/2$. Because $y \varepsilon U$, $\rho(x, y) < \epsilon/2$. Hence, since $x = h(x, 0)$, $\rho(h(x, 0), h(y, t)) < \epsilon/2 + \epsilon/2 = \epsilon$. Thus h is continuous at $(x, 0)$, and therefore is continuous.

We have shown that $h: K \times I \to K$ is a homotopy.[10] Since $h(x, 0) = x$, and $h(x, 1) = fg_1(x)$, we have

(15.1) *The map $fg_1: K \to K$ is homotopic to the identity.*

We have in fact proved much more. We have shown that if $0 \leq t \leq 1/n$, then $\rho(y, h(y, t)) < 1/n$. Hence if $0 \leq t \leq 1$, $\rho(y, h(y, t/n)) < 1/n$. Thus if we set $h_n(x, t) = h(x, t/n)$, we have $h_n(x, 0) = h(x, 0) = x$, and $h_n(x, 1) = h(x, 1/n) = \psi_n\phi_n(x)$, and therefore h_n is a homotopy of the identity

[10] Actually the homotopy is a uniform homotopy. For definition of uniform homotopy see [3], p. 204.

map of K in K to the map $\psi_n\phi_n$ such that, for each $x \in K$ and $t \in I$, $\rho(x, h_n(x,t)) < 1/n$. Thus if we are given $\epsilon > 0$, and if we choose $n > 1/\epsilon$, then for every $t \in I$, $h_n(x,t)$ is within ϵ of x. We have therefore obtained

(15.2)[11] *Given any positive number ϵ, the identity map of the metric complex K on itself is ϵ-homotopic to a factored map $\psi_n\phi_n$, where ϕ_n is a map of K into a Whitehead complex N^n, and ψ_n is a map of N^n into K. During the homotopy a point x does not leave the closure $\bar{e}(x)$ of the cell containing it.*

16. The homotopy type. Two spaces X and Y are said to have the same *homotopy type* if there exist maps $f: X \to Y$ and $g: Y \to X$ such that $fg: Y \to Y$ is homotopic to the identity, and $gf: X \to X$ is homotopic to the identity. Spaces of the same homotopy type can not be distinguished by any of the invariants of algebraic topology. Our main theorem is that a metric complex and the corresponding Whitehead complex have the same homotopy type.

THEOREM 1. *If K is a metric*[12] *complex and K_W is the complex retopologized with the Whitehead topology, then K and K_W have the same homotopy type.*

Proof. We have maps $f: K_W \to K$ and $g_1: K \to K_W$ such that, by (15.1), $fg_1: K \to K$ is homotopic to the identity. It is then sufficient to show that $g_1f: K_W \to K_W$ is homotopic to the identity. We define the homotopy $h: K_W \times I \to K_W$ as follows. For each $x \in K_W$, $g_1f(x) = g_1(x) \in \bar{e}(x)$; we define $h(x,t)$ to be the point which divides the segment from x to $g_1f(x)$ in $\bar{e}(x)$ in the ratio $t:1-t$. For each cell $\bar{e}_a$ of K_W, if $x \in \bar{e}_a$, then $e(x) \leq e_a$, and hence $h(x,t)$ is the point dividing the segment from x to $g_1f(x)$ in $\bar{e}_a$ in the ratio $t:(1-t)$. Thus $h \mid \bar{e}_a \times I$ is continuous, and by (5.2), $h: K_W \times I \to K_W$ is a homotopy.

THEOREM 2. *Isomorphic metric complexes have the same homotopy type.*

Proof. Let K and L be isomorphic metric complexes. Then K_W and L_W are isomorphic Whitehead complexes. Hence by (6.2), K_W and L_W are homeomorphic and a fortiori of the same homotopy type. By Theorem 1,

---

[11] In the terminology of Lefschetz ([9], p. 98), (15.2) says that the identity mapping of K on itself is ϵ-deformable, for all ϵ, to a mapping into a continuous complex. We may interpret this as meaning that K is an absolute neighborhood retract.

[12] In this and the following theorems it is sufficient to assume that K is a metrizable complex.

K has the same homotopy type as K_W, and L has the same homotopy type as L_W. Hence K has the same homotopy type as L.

17. Canonical mappings. We now show that the theorems on canonical mappings [13] into the nerve of a covering hold equally whether the nerve is provided with a metric or with the Whitehead topology.

THEOREM 3. *Let X be a topological space, let $\mathfrak{U}$ be a covering of X, let N be the nerve of $\mathfrak{U}$ with the Whitehead topology, and let M be the nerve metrized in some way to form a metric complex. Then there exists a canonical map of X into the nerve N of $\mathfrak{U}$ if and only if there is a canonical map of X into the nerve M of $\mathfrak{U}$.*

Proof. First let ϕ be a canonical map of X into N, and let f be the identity map of N $(= M_W)$ onto M. Then $f\phi : X \to M$ is continuous. Also, for each vertex u_λ of M, $(f\phi)^{-1}\mathrm{St}_M(u_\lambda) = \phi^{-1}\mathrm{St}_N(u_\lambda) \subset U_\lambda$. Hence $f\phi$ is a canonical map of X into M.

On the other hand, let ψ be a canonical map of X into M. Let g be a map of M into N $(= M_W)$ (see § 14) such that, for each $x \in M$, $g(x) \in \delta(x)$ in N. Then $g\psi$ is a map of X into N. If $g(x) \in e_a$, $x \in \mathrm{St}\, e_a$; hence $g^{-1}e_a \subset \mathrm{St}\, e_a$. If $e_\gamma \subset \mathrm{St}\, e_a$, $g^{-1}e_\gamma \subset \mathrm{St}\, e_\gamma \subset \mathrm{St}\, e_a$. Hence $g^{-1}\mathrm{St}_N e_a \subset \mathrm{St}_M e_a$. It follows that $(g\psi)^{-1}\,\mathrm{St}_N\, u_\lambda = \psi^{-1}g^{-1}\,\mathrm{St}_N\, u_\lambda \subset \psi^{-1}\,\mathrm{St}_M\, u_\lambda \subset U_\lambda$. Therefore $g\psi$ is a canonical map of X into N.

COROLLARY 1. *Let X be a normal space, and let $\mathfrak{U}$ be a covering of X. Then there is a canonical map of X into a metric nerve M of $\mathfrak{U}$, or into the Whitehead nerve N of $\mathfrak{U}$, if and only if $\mathfrak{U}$ has a locally finite refinement.*

COROLLARY 2. *Let X be a topological space. There is a canonical map of X into a metric nerve M or into the Whitehead nerve N of every covering $\mathfrak{U}$ of X if and only if X is paracompact and normal.*

Proof. These results have been proved for particular metric nerves ([4], p. 388). It follows from Theorem 3 that they hold for Whitehead nerves. By another application of Theorem 3 they hold for any other metric nerves.

18. Topological invariance. Finally we show the topological invariance invariance of the homology and cohomology groups of metric complexes.

THEOREM 4. *The combinatorial homology and cohomology groups of a*

---

[13] For definition of canonical mappings see section 7 above.

metric complex are isomorphic with the corresponding singular homology groups. If the coefficient group is discrete, the combinatorial cohomology groups are isomorphic with the Čech cohomology groups.

Proof. The singular homology and cohomology groups and the Čech cohomology groups are invariants of the homotopy type ([7], p. 400; [5], p. 287). Hence by Theorem 1 they are the same for K and for K_W. By (8.1) and (8.2) they are therefore isomorphic with the corresponding combinatorial homology and cohomology groups.

COROLLARY. *If K and L are homeomorphic metric complexes, their combinatorial homology and cohomology groups are isomorphic.*

HARVARD UNIVERSITY.

BIBLIOGRAPHY

[1] P. Alexandroff and H. Hopf, *Topologie*, Berlin (1935).

[2] N. Bourbaki, *Éléments de Mathematique* II, (Actualités scientifiques et industrielles, No. 858), Paris (1940).

[3] C. H. Dowker, "Mapping theorems for non-compact spaces," *American Journal of Mathematics*, vol. 69 (1947), pp. 200-242.

[4] ———, "An extension of Alexandroff's mapping theorem," *Bulletin of the American Mathematical Society*, vol. 54 (1948), pp. 386-391.

[5] ———, "Čech cohomology theory and the axioms," *Annals of Mathematics*, vol. 51 (1950), pp. 276-292.

[6] S. Eilenberg, "Singular homology theory," *Annals of Mathematics*, vol. 45 (1944), pp. 407-447.

[7] W. Hurewicz, J. Dugundji and C. H. Dowker, "Continuous connectivity groups in terms of limit groups," *Annals of Mathematics*, vol. 49 (1948), pp. 391-406.

[8] S. Lefschetz, *Algebraic Topology*, New York (1942).

[9] ———, *Topics in Topology*, Princeton (1942).

[10] K. Reidemeister, *Topologie der Polyeder*, Leipzig (1938).

[11] J. H. C. Whitehead, "Simplicial spaces, nuclei and m-groups," *Proceedings of the London Mathematical Society* (2), vol. 45 (1939), pp. 243-327.

[12] ———, "Combinatorial homotopy I," *Bulletin of the American Mathematical Society*, vol. 55 (1949), pp. 213-245.

ON THE UNBOUNDEDNESS OF THE ESSENTIAL SPECTRUM.*

By C. R. Putnam.

1. In the differential equation

$$x'' + (\lambda - f)x = 0, \tag{1}$$

let λ denote a real parameter, and let $f = f(t)$ be a real-valued continuous function on the half-line $0 \leqq t < \infty$. (Throughout this paper only real-valued functions will be considered.) In addition, suppose that f is such that the differential equation (1) is of the Grenzpunkt type, so that (1) possesses for some λ (and hence for every λ) a solution x which fails to belong to the class $L^2 = L^2[0, \infty)$; *cf.* Weyl [11], p. 238. In this case, the equation (1) and a linear, homogeneous boundary condition

$$x(0) \cos \alpha + x'(0) \sin \alpha = 0, \qquad 0 \leqq \alpha < \pi, \tag{2}$$

determine a boundary value problem on $0 \leqq t < \infty$ with a spectrum S_α. If S' denotes the (closed, possibly empty) set of cluster points of S_α, then S' is independent of α ([11], p. 251), and is called the essential spectrum of (1).

Various results concerning the set S' are known in case f is subject to certain restrictions. If, for instance, f satisfies

$$f(t) \to c \quad \text{as } t \to \infty, \tag{3}$$

then S' is the half-line $c \leqq t < \infty$; [3]. In general, the complement of S' is an open (possibly empty) set, and hence possesses a decomposition $\Sigma(\lambda_k, \lambda^k)$ into open intervals $\lambda_k < \lambda < \lambda^k$ (or "gaps" in S'), where it is understood that one, or possibly two, of the intervals are half-lines, and that the summation may consist of the single interval $-\infty < \lambda < \infty$ in case S' is empty. If f is bounded, so that

$$|f(t)| < \text{const.}, \qquad 0 \leqq t < \infty, \tag{4}$$

it is known that, except for $(\lambda_0, \lambda^0) = (-\infty, \lambda^0)$, the inequality $\lambda^k - \lambda_k \leqq$ const. holds, so that, in particular, S' is unbounded from above ([8]), and if certain extra conditions are placed on f, even asymptotic estimates of the gaps in S' can be given; *cf.* [6].

* Received September 15, 1951.

metric complex are isomorphic with the corresponding singular homology groups. If the coefficient group is discrete, the combinatorial cohomology groups are isomorphic with the Čech cohomology groups.

Proof. The singular homology and cohomology groups and the Čech cohomology groups are invariants of the homotopy type ([7], p. 400; [5], p. 287). Hence by Theorem 1 they are the same for K and for K_W. By (8.1) and (8.2) they are therefore isomorphic with the corresponding combinatorial homology and cohomology groups.

COROLLARY. *If K and L are homeomorphic metric complexes, their combinatorial homology and cohomology groups are isomorphic.*

HARVARD UNIVERSITY.

BIBLIOGRAPHY

[1] P. Alexandroff and H. Hopf, *Topologie*, Berlin (1935).

[2] N. Bourbaki, *Éléments de Mathematique* II, (Actualités scientifiques et industrielles, No. 858), Paris (1940).

[3] C. H. Dowker, "Mapping theorems for non-compact spaces," *American Journal of Mathematics*, vol. 69 (1947), pp. 200-242.

[4] ———, "An extension of Alexandroff's mapping theorem," *Bulletin of the American Mathematical Society*, vol. 54 (1948), pp. 386-391.

[5] ———, "Čech cohomology theory and the axioms," *Annals of Mathematics*, vol. 51 (1950), pp. 276-292.

[6] S. Eilenberg, "Singular homology theory," *Annals of Mathematics*, vol. 45 (1944), pp. 407-447.

[7] W. Hurewicz, J. Dugundji and C. H. Dowker, "Continuous connectivity groups in terms of limit groups," *Annals of Mathematics*, vol. 49 (1948), pp. 391-406.

[8] S. Lefschetz, *Algebraic Topology*, New York (1942).

[9] ———, *Topics in Topology*, Princeton (1942).

[10] K. Reidemeister, *Topologie der Polyeder*, Leipzig (1938).

[11] J. H. C. Whitehead, "Simplicial spaces, nuclei and *m*-groups," *Proceedings of the London Mathematical Society* (2), vol. 45 (1939), pp. 243-327.

[12] ———, "Combinatorial homotopy I," *Bulletin of the American Mathematical Society*, vol. 55 (1949), pp. 213-245.

ON THE UNBOUNDEDNESS OF THE ESSENTIAL SPECTRUM.*

By C. R. PUTNAM.

1. In the differential equation

$$x'' + (\lambda - f)x = 0, \tag{1}$$

let λ denote a real parameter, and let $f = f(t)$ be a real-valued continuous function on the half-line $0 \leqq t < \infty$. (Throughout this paper only real-valued functions will be considered.) In addition, suppose that f is such that the differential equation (1) is of the Grenzpunkt type, so that (1) possesses for some λ (and hence for every λ) a solution x which fails to belong to the class $L^2 = L^2[0, \infty)$; *cf.* Weyl [11], p. 238. In this case, the equation (1) and a linear, homogeneous boundary condition

$$x(0) \cos \alpha + x'(0) \sin \alpha = 0, \qquad 0 \leqq \alpha < \pi, \tag{2}$$

determine a boundary value problem on $0 \leqq t < \infty$ with a spectrum S_α. If S' denotes the (closed, possibly empty) set of cluster points of S_α, then S' is independent of α ([11], p. 251), and is called the essential spectrum of (1).

Various results concerning the set S' are known in case f is subject to certain restrictions. If, for instance, f satisfies

$$f(t) \to c \quad \text{as } t \to \infty, \tag{3}$$

then S' is the half-line $c \leqq t < \infty$; [3]. In general, the complement of S' is an open (possibly empty) set, and hence possesses a decomposition $\Sigma(\lambda_k, \lambda^k)$ into open intervals $\lambda_k < \lambda < \lambda^k$ (or "gaps" in S'), where it is understood that one, or possibly two, of the intervals are half-lines, and that the summation may consist of the single interval $-\infty < \lambda < \infty$ in case S' is empty. If f is bounded, so that

$$|f(t)| < \text{const.}, \qquad 0 \leqq t < \infty, \tag{4}$$

it is known that, except for $(\lambda_0, \lambda^0) = (-\infty, \lambda^0)$, the inequality $\lambda^k - \lambda_k \leqq$ const. holds, so that, in particular, S' is unbounded from above ([8]), and if certain extra conditions are placed on f, even asymptotic estimates of the gaps in S' can be given; *cf.* [6].

* Received September 15, 1951.

In most instances, the set S' is, if it is not empty, unbounded either from above or from below. However, examples of functions f for which S' consists of a single point are known; [4], Corollary 3, p. 110. Thus the set S' can be non-empty and yet bounded.

It is natural to ask under what general condition on f is the set S' unbounded when it is not empty. It turns out that a sufficient condition is that f be bounded from below, thus, f should satisfy

$$f(t) > \text{const.}, \qquad 0 \leqq t < \infty. \tag{5}$$

(It should be noticed that (5), and hence (4) or (3), imply that (1) is of the Grenzpunkt type; [11], p. 238.) Furthermore, as will be shown in Theorem I below, the restriction (5) even permits an asymptotic estimate of the gaps in S'. For convenience in the sequel, the following terminology will be introduced: Let $m_\alpha(\lambda) = \min |\lambda - \mu|$ when μ is in S_α, and let $m(\lambda) = \min |\lambda - \mu|$ when μ is in S'. Thus $m_\alpha(\lambda)$ and $m(\lambda)$ denote the distance from a fixed value λ to the nearest point of the set S_α or S'. Clearly $m(\lambda) = \infty$ for some λ (and hence for every λ) if and only if S' is empty. The following will be proved:

THEOREM I. *Let $f(t)$ be a real-valued continuous function on $0 \leqq t < \infty$ satisfying condition* (5). *Then one of the following two possibilities must occur: either* (i) *the set S' is empty or* (ii) *S' is not empty and is unbounded from above. Furthermore, in case* (ii), *$m(\lambda)$ satisfies*

$$m(\lambda) = O(\lambda^{\frac{1}{2}}), \qquad \lambda \to \infty. \tag{6}$$

The main tool used in this paper will be the lemma (*) of section **2** from which certain estimates for $m(\lambda)$ will be derived; various consequences of (*), in addition to Theorem I, will be set forth in Theorems II-IV in section **4**.

2. Consider the boundary value problem determined by (1) and (2) for a fixed α, and let $\lambda_1, \lambda_2, \cdots$ denote the eigenvalues (if any) and $\phi_1, \phi_2, \cdots$ the corresponding normalized eigenfunctions. Then for $\lambda = \lambda_j$ and $x = \phi_j$, equation (1) becomes

$$\phi_j'' + (\lambda_j - f)\phi_j = 0. \tag{7}$$

Let $L(x)$ be defined by $L(x) = x'' - fx$, and let g denote any function of class L^2 satisfying the boundary condition (2), for which $L(g)$ is defined, continuous, and of class L^2; thus,

$$\int_0^\infty g^2 dt < \infty, \qquad \int_0^\infty (L(g))^2 dt < \infty. \tag{8}$$

Since g satisfies (2), $\phi_j'(0)g(0) - \phi_j(0)g'(0) = 0$; in addition, by (8), $(\phi_j'(t)g(t) - \phi_j(t)g'(t)) \to 0$ as $t \to \infty$ (cf. [11], pp. 241-242). Multiplication of (7) by g followed by an integration readily leads to

$$(9) \qquad \int_0^\infty (L(g) + \lambda g)\phi_j dt = (\lambda - \lambda_j) \int_0^\infty g\phi_j dt,$$

for an arbitrary real number λ. Two applications of the Parseval relation applied to the functions $L(g) + \lambda g$ and g yield, in virtue of (9) and a similar relation in which the ϕ_j are replaced by the eigendifferentials corresponding to the continuous spectrum, the inequality

$$(10) \qquad \int_0^\infty (L(g) + \lambda g)^2 dt \geqq m_\alpha^2(\lambda) \int_0^\infty g^2 dt;$$

cf. [9], p. 140, for calculations of a similar nature.

First suppose that S' is not empty, so that $m(\lambda) < \infty$ (for all λ), and suppose $m_\alpha(\lambda) < m(\lambda)$. Let $g = g_n$, $n = 1, 2, \cdots$, denote any sequence of functions of the type considered above in the derivation of (10). In addition, suppose that

$$(11) \qquad \int_0^T g_n^2 dt \Big/ \int_0^\infty g_n^2 dt \to 0, \qquad n \to \infty,$$

holds for every fixed T satisfying $0 \leqq T < \infty$. Let ϵ denote a small positive number, and consider the λ-interval $[\lambda - m(\lambda) + \epsilon, \lambda + m(\lambda) - \epsilon]$. On this interval there exist at most a finite number of points $\lambda_1, \lambda_2, \cdots, \lambda_N$ (eigenvalues) in the spectrum of S_α. It follows from the Schwarz inequality and (11) that

$$(12) \qquad \sum_{j=1}^N \left(\int_0^\infty g_n\phi_j dt\right)^2 \Big/ \int_0^\infty g_n^2 dt \to 0, \qquad n \to \infty.$$

It is now an easy consequence of (12) that

$$\limsup_{n\to\infty} \int_0^\infty (L(g_n) + \lambda g_n^2) dt \geqq \limsup_{n\to\infty} (m(\lambda) - \epsilon)^2 \int_0^\infty g_n^2 dt,$$

whenever $\epsilon < m(\lambda)$ and the functions $g = g_n$ satisfy (11). Since ϵ is arbitrary, one can obtain even

$$(13) \qquad \limsup_{n\to\infty} \int_0^\infty (L(g_n) + \lambda g_n)^2 dt \geqq \limsup_{n\to\infty} [m^2(\lambda) \int_0^\infty g_n^2 dt].$$

Next, let y denote any function satisfying (2) for which y, y', and $L(y)$ are continuous and belong to L^2. Furthermore, let $\lambda > 0$, and put $h = \cos(\lambda^{\frac{1}{2}} t)$ and $g = yh$. It is readily verified that $g \cos \alpha + g' \sin \alpha = h(y \cos \alpha + y' \sin \alpha)$

$+ yh' \sin \alpha$, so that, since y satisfies (2) and since $h' = -\lambda^{\frac{1}{2}} \sin(\lambda^{\frac{1}{2}} t)$, g clearly satisfies (2). Furthermore, $L(g) + \lambda g = L(y)h + 2y'h'$, so that the left side of (10) becomes

$$\int_0^\infty (L(y)h + 2y'h')^2 dt. \tag{14}$$

One readily verifies as a consequence of the equation

$$\cos^2(\lambda^{\frac{1}{2}} t) = \tfrac{1}{2}(1 + \cos(2\lambda^{\frac{1}{2}} t))$$

and an integration by parts, that the integral on the right side of (10) is

$$\int_0^\infty y^2 h^2 dt = \tfrac{1}{2}\int_0^\infty y^2 dt - \tfrac{1}{2}\lambda^{-\frac{1}{2}} \int_0^\infty yy' \sin(2\lambda^{\frac{1}{2}} t)\, dt. \tag{15}$$

(Use is made of the limit relation $y(t) \to 0$ as $t \to \infty$; this, in turn, is a consequence of the fact that y and y' belong to L^2.) An application of the Schwarz inequality to the second integral on the right side of equation (15) shows that

$$\int_0^\infty yy' \sin(2\lambda^{\frac{1}{2}} t)\, dt \leqq \left(\int_0^\infty y^2 dt \int_0^\infty y'^2 dt\right)^{\frac{1}{2}}. \tag{16}$$

If use is made of the inequality $(a + b)^2 \leqq 2(a^2 + b^2)$, where a and b are real, it is seen that (14) is majorized by

$$2\int_0^\infty [(L(y))^2 + 4\lambda y'^2]\, dt. \tag{17}$$

Furthermore, if y is normalized by

$$\int_0^\infty y^2 dt = 1, \tag{18}$$

it is seen from (15) and (16) that

$$\int_0^\infty y^2 h^2 dt \geqq \tfrac{1}{2} - \tfrac{1}{2}\lambda^{-\frac{1}{2}} \left(\int_0^\infty y'^2 dt\right)^{\frac{1}{2}}. \tag{19}$$

Suppose now that $y = y_n$, $n = 1, 2, \cdots$, denotes any sequence of functions, satisfying the conditions imposed on y above, and such that (11), in which $g_n = y_n h$, is valid for every fixed T, where $0 \leqq T < \infty$. It then follows from (13) and the results obtained above that

$$\begin{aligned} (20)\qquad & 4 \limsup_{n\to\infty} \int_0^\infty [(L(y_n))^2 + 4\lambda y_n'^2]\, dt \\ & \qquad \geqq m^2(\lambda)\left[1 - \lambda^{-\frac{1}{2}} \liminf_{n\to\infty} \left(\int_0^\infty y_n'^2 dt\right)^{\frac{1}{2}}\right]. \end{aligned}$$

It is clear from (19) that the condition (11) surely holds if

$$(21)\qquad 0 < \int_0^T y_n^2 dt / [1 - \lambda^{-\frac{1}{2}} (\int_0^\infty y_n'^2 dt)^{\frac{1}{2}}] \to 0, \qquad n \to \infty.$$

Furthermore, it is clear from the above discussion, that (20) is also valid in case S' is empty (so that $m(\lambda) = \infty$), with the understanding that the left side of the inequality of (20) may be ∞. The results obtained thus far can be summarized in the following lemma:

(*) *Let* $\{y_n\}$ *denote any sequence of real-valued functions on* $0 \leqq t < \infty$, *satisfying the boundary condition* (2) *for* $x = y_n$ *and the normalization condition* (18) *for* $y = y_n$, *and which, in addition, are such that* y_n, y'_n *and* $L(y_n)$ *are continuous, belong to* L^2, *and the inequality and limit relation of* (21) *are satisfied. Then the inequality* (20) *is valid for all* $\lambda > 0$ (*where* $m(\lambda) \leqq \infty$).

The above lemma will be used in the next section to prove Theorem I.

3. *Proof of Theorem* I. It is sufficient to show in this case that if S' is not empty, then (6) must hold. For it is an obvious consequence of (6) that S' must contain an infinity of points clustering at $\lambda = +\infty$ (and possibly elsewhere). But if S' is not empty, then S' is translated by c if $f(t)$ is replaced by $f(t) + c$, for any constant c. Since (5) remains valid for $f(t) + c$ whenever it holds for $f(t)$, it can be supposed that $\lambda = 0$ belongs to the set S'. Consider the fixed boundary value problem determined by (1) and (2) for $\alpha = 0$. Suppose first that $\lambda = 0$ is a cluster point of eigenvalues λ_n with corresponding normalized eigenfunctions ϕ_n, so that

$$(22)\qquad \int_0^\infty \phi_n^2 dt = 1.$$

It is clear that

$$(23)\qquad \int_0^\infty (L(\phi_n))^2 dt \to 0, \qquad n \to \infty.$$

Furthermire, multiplication of the equation (7) by ϕ_n for $j = n$, followed by an integration, yields

$$(24)\qquad \int_0^T \phi_n'^2 dt = \phi_n \phi_n' \Big|_0^T + \int_0^T (\lambda_n - f) \phi_n^2 dt,$$

for every T satisfying $0 \leqq T < \infty$. However, it follows from (5) that $-f <$ const., so that, since $\phi_n(0) = 0$,

$$\int_0^T \phi_n'^2 dt \leqq \phi_n(T) \phi_n'(T) + \text{const.}$$

It is known that $\phi_n(T)\phi_n'(T) \to 0$ as $T \to \infty$ (see the proof of the theorem of [12], p. 6; cf. also [2]) so that

$$(25) \qquad \int_0^\infty \phi_n'^2 dt \leqq \text{const.} \qquad \text{(independent of } n\text{)}.$$

Furthermore, the limit relation $\phi_n(t) \to 0$ as $n \to \infty$ holds uniformly on every finite t-interval $0 \leqq t \leqq T$; cf. [10], p. 269.[1] Hence, for sufficiently large λ, it follows from (25), that (21) holds for $y_n = \phi_n$. Moreover, it can be supposed that (for sufficiently large λ),

$$1 - \lambda^{-\frac{1}{2}} \limsup_{n\to\infty} \left(\int_0^\infty y_n'^2 dt \right)^{\frac{1}{2}} \geqq \tfrac{1}{2}$$

The lemma (*) is clearly applicable and shows that (20) holds and hence, by (22)-(24),

$$(26) \qquad \text{const.}\, \lambda \geqq \tfrac{1}{2} m^2(\lambda)$$

for large λ. That is, (6) is satisfied and the proof of Theorem I is complete, at least if $\lambda = 0$ is a cluster point of eigenvalues of the boundary value problem corresponding to $\alpha = 0$. If this last condition fails to hold, then $\lambda = 0$ is in the continuous spectrum of this boundary problem. (This alone would imply that S' is unbounded; [13].) It will be shown that again (6) must hold. (The proof is essentially similar to that carried out in case $\lambda = 0$ is a cluster point of eigenvalues.) One can obtain a sequence of functions y_n of the type considered in (*), so that, in particular, (25) holds if ϕ_n is replaced by y_n, and such that

$$(27) \qquad y_n'' + (\lambda_n - f) y_n = k_n,$$

where the k_n are continuous functions satisfying

$$(28) \qquad \int_0^\infty k_n^2 dt \to 0, \qquad n \to \infty;$$

see the lemma of [10], p. 269. In addition (23) is valid, and the limit relation $y_n(t) \to 0$, as $n \to \infty$ holds uniformly on every fixed t-interval $0 \leqq t \leqq T$; *loc. cit.*, p. 269. Finally, a relation of the type (25), but where ϕ_n is replaced by y_n, follows from (5), (27) and (28). Hence (26) can again be obtained and the proof of Theorem I is complete.

---

[1] Professor Wintner has pointed out to me that the passage occurring in [10], p. 267, referring to a classical theorem in [1], p. 278, should state that the methods of [1] imply that the class of an operator, on the L^2 space $0 \leqq t < \infty$, is unchanged by adding to it a bounded operator. Cf. a corresponding remark in [7], § 7.

4. In this section, a number of additional consequences of the lemma (*) of section **2** will be derived. First, a slight generalization of Theorem I is contained in the following

THEOREM II. *The essential spectrum S' is not empty and, in fact, holds when the assumption that f satisfies* (5) *is replaced by the assumption that* (1) *be of the Grenzpunkt type and that there exist a sequence of real-valued functions y_n satisfying the following three conditions:*

(i) *the y_n possess continuous second derivatives and satisfy the boundary condition* (2) *for a fixed α;*

(ii) $\int_0^\infty y_n^2 dt = 1$, $\int_0^T y_n^2 dt \to 0$ *(as $n \to \infty$, for every fixed positive number T);*

(iii) $\int_0^\infty y_n'^2 dt <$ const., $\int_0^\infty (L(y_n))^2 dt <$ const.

That Theorem II implies Theorem I is clear. In fact, the conditions specified in II were obtained, in the proof of I, as a consequence of the assumption that (5) holds and that S' is not empty. The proof of III proceeds along the same lines as that of I, and can therefore be omitted.

THEOREM III. *If, in Theorem* II, *the first inequality in condition* (iii) *is strengthened to*

$$\int_0^\infty y_n'^2 dt \to 0, \qquad n \to \infty, \tag{29}$$

then the assertion (6) *of Theorem* II *can be improved to the statement*

$$m(\lambda) = O(1), \quad \text{as } \lambda \to \infty. \tag{30}$$

If, in addition to (29), *the second inequality of condition* (iii) *is replaced by*

$$\int_0^\infty (L(y_n))^2 dt \to 0, \qquad n \to \infty, \tag{31}$$

then the assertion of Theorem II *can be improved* to

$$\text{the half-line } 0 \leqq \lambda < \infty \text{ is contained in } S'. \tag{32}$$

In order to prove Theorem III one need note only that (20) now yields

$$\text{const.} \geqq m^2(\lambda) \qquad \text{or} \quad 0 \geqq m^2(\lambda), \tag{33}$$

according as (29) alone or both (29) and (31) are assumed. Since (33) is equivalent to (30) or (32), the proof of Theorem III is complete.

THEOREM IV. *Let $f(t)$ be a real-valued continuous function on $0 \leqq t < \infty$ satisfying (5), and suppose that, as $t \to \infty$, the (finite) value $\mu = \liminf f(t)$ belongs to the set S'. Then the set S' is precisely the half-line $\mu \leqq \lambda < \infty$.*

That S' is contained in the half-line $\mu \leqq \lambda < \infty$ is a consequence of the fact that the least point of S' is never less than μ; cf. [5], p. 850. It remains to show then that every $\lambda > \mu$ (hence every $\lambda \geqq \mu$) is in S'. To this end, as in the proof of Theorem I, it can be supposed that $f(t)$ is shifted, if necessary, by a constant, so that $\mu = 0$. Let $\lambda > 0$. It will be shown that there exists a sequence of functions y_n satisfying the conditions of Theorem III sufficient for the implication (32). For convenience it will be supposed that $\alpha = 0$ in (2) and that $\mu = 0$ is a cluster point of eigenvalues. (Note that the set S' is independent of α; moreover, the case in which $\mu = 0$ is not a cluster point of eigenvalues, and hence is in the continuous spectrum, can be treated as in the proof of Theorem I.) Again, one can obtain relation (23). It is clear from (24), the fact that $\mu = 0$, and from the properties of the functions ϕ_n occurring in the proof of Theorem I that

$$\limsup \int_0^\infty \phi_n'^2 dt = \limsup \left(-\int_0^\infty f\phi_n^2 dt\right) \leqq 0, \qquad (n \to \infty),$$

as an improvement to (25). The last formula line, for $y_n = \phi_n$, of course implies (29). It is now clear that the functions $y_n = \phi_n$ satisfy the conditions of Theorem III guaranteeing the validity of (32). This completes the proof of Theorem IV.

PURDUE UNIVERSITY.

REFERENCES.

[1] T. Carleman, *Sur les équations intégrales singulières à noyau réel et symétrique*, Uppsala, 1923.

[2] P. Hartman, "The L^2-solutions of linear differential equations of second order," *Duke Mathematical Journal*, vol. 14 (1947), pp. 323-326.

[3] ———, "On the spectra of slightly disturbed linear oscillators," *American Journal of Mathematics*, vol. 71 (1949), pp. 71-79.

[4] ———, "Some examples in the theory of singular boundary value problems," *ibid.*, *vol.* 74 (1952), pp. 107-126.

[5] ——— and C. R. Putnam, "The least cluster point of the spectrum of boundary value problems," *ibid.*, vol. 70 (1948), pp. 849-855.

[6] ——— and C. R. Putnam, "The gaps in the essential spectra of wave equations," *ibid.*, vol. 72 (1950), pp. 849-862.

[7] ——— and A. Wintner, "On perturbations of the continuous spectrum of the harmonic oscillator," *ibid.*, vol. 74 (1952), pp. 79-85.

[8] C. R. Putnam, "The cluster spectra of bounded potentials," *ibid.*, vol. 70 (1949), pp. 842-848.

[9] ———, "On isolated eigenfunctions associated with bounded potentials," *ibid.*, vol. 72 (1950), pp. 135-147.

[10] ———, "The comparison of spectra belonging to potentials with a bounded difference," *Duke Mathematical Journal*, vol. 18 (1951), pp. 267-273.

[11] H. Weyl, "Ueber gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen," *Mathematische Annalen*, vol. 68 (1910), pp. 222-269.

[12] A. Wintner, "(L^2)-connections between the potential and kinetic energies of linear systems," *American Journal of Mathematics*, vol. 72 (1947), pp. 5-13.

[13] ———, "On Dirac's theory of continuous spectra," *The Physical Review*, vol. 73 (1948), pp. 781-785.

PROPERTIES OF CONFORMAL INVARIANTS.*

By VIDAR WOLONTIS.**

I. Basic Properties of Extremal Distance.

1 **Definition of extremal distance.**[1] Let D be a region in the complex plane, E_1 and E_2 two disjoint compact subsets of D, and Γ the class of all rectifiable curves in D joining E_1 and E_2. Let P be the set of non-negative functions $\rho(z)$ on D such that the integral

$$L_\rho(\gamma) = \int_\gamma \rho \mid dz \mid \tag{1}$$

is defined for all $\rho \varepsilon \mathrm{P}$ and every rectifiable curve γ in D, and

$$A_\rho(D) = \int\int_D \rho^2 dxdy, \qquad z = x + iy, \tag{2}$$

exists and is different from zero. We wish to determine $\rho \varepsilon \mathrm{P}$ so that the ratio

$$[\inf_{\gamma \varepsilon \Gamma} L_\rho(\gamma)]^2 / A_\rho(D) \tag{3}$$

be maximal. Since the existence of a maximum is not assured, we consider more generally the finite or positively infinite quantity

$$\lambda_D(E_1, E_2) = \sup_{\rho \varepsilon \mathrm{P}} [\inf_{\gamma \varepsilon \Gamma} L_\rho(\gamma)]^2 / A_\rho(D), \tag{4}$$

which we call the *extremal distance* between E_1 and E_2 with respect to the region D. For later use we observe that, since the value of the quotient (3) remains unchanged if ρ is multiplied by a positive constant, and since the extremal distance (4) is clearly never zero, we may restrict the class P e. g. by

* Received July 10, 1951.

** This paper includes the results found in my thesis, "Properties of Conformal Invariants," Harvard University, 1949. I wish to express my deep gratitude to Professor Lars V. Ahlfors for suggesting problems and methods, and for encouraging guidance and great personal interest.

[1] For the material presented in sections 1-3 of this chapter I am indebted to Professors Ahlfors and Beurling for their permission to consult an unpublished manuscript. Compare also A. Beurling, *Etudes sur un problème de majoration*, Thèse, Uppsala, 1933, and L. Ahlfors and A. Beurling, "Conformal invariants and function-theoretic null-sets," *Acta Mathematica*, vol. 83 (1950), pp. 101-129.

the requirement $L_\rho(\gamma) \geq 1$, for all ρ and γ, in which case (4) takes the simple form

$$(4') \qquad \lambda_D(E_1, E_2) = \sup_\rho 1/A_\rho(D), \qquad L_\rho(\gamma) \geq 1.$$

The extremal distance is a *conformal invariant* of the configuration (D, E_1, E_2), i. e. if $z^* = f(z)$ is a one-to-one conformal mapping of D upon a region D^*, taking E_1 to E^*_1 and E_2 to E^*_2, then

$$(5) \qquad \lambda_{D^*}(E^*_1, E^*_2) = \lambda_D(E_1, E_2).$$

In fact, with $\rho^*(z^*) = \rho(z)/|f'(z)|$ we have

$$L_{\rho^*}(\gamma^*) = \int_{\gamma^*} \rho^* \, | dz^* | = \int_\gamma \rho \, |dz| = L_\rho(\gamma),$$

$$A_{\rho^*}(D^*) = \int\int_{D^*} \rho^{*2} dx^* dy^* = \int\int_D \rho^2/|f'(z)|^2 dx^* dy^*$$

$$= \int\int_D \rho^2 dx dy = A_\rho(D).$$

The definition of extremal distance remains meaningful if we allow E_1 and E_2 to contain accessible boundary points of D.

2. The extension principle. The following simple principle is an important tool in dealing with extremal distances: If D^* is a region containing D, and if E_1^* and E_2^* are compact subsets of D^* containing E_1 and E_2 respectively, then

$$(6) \qquad \lambda_{D^*}(E_1^*, E_2^*) \leq \lambda_D(E_1, E_2).$$

The proof is immediate: Any $\rho(z)$ defined on D^* is also defined on D,

$$(7) \qquad A_\rho(D) \leq A_\rho(D^*),$$

and, since any curve joining E_1 and E_2 will a fortiori belong to the class Γ^* of curves joining E_1^* and E_2^*,

$$(8) \qquad \inf L_\rho(\gamma) \geq \inf L_\rho(\gamma^*), \text{ where } \gamma \, \varepsilon \, \Gamma, \ \gamma^* \, \varepsilon \, \Gamma^*.$$

Inserted into the definition (4), these inequalities yield (6).

In this connection we observe that if E_1' and E_2' denote the boundaries [2] of E_1 and E_2, we have $E_1' \subset E_1$, $E_2' \subset E_2$; and (6) gives

$$(9) \qquad \lambda_D(E_1, E_2) \leq \lambda_D(E_1', E_2').$$

---

[2] Throughout this paper primed letters will denote the boundaries of the respective sets.

On the other hand every curve joining E_1 and E_2 will also join their boundaries, which implies, in analogy with (8), that

$$\lambda_D(E_1, E_2) \geq \lambda_D(E_1', E_2'). \tag{10}$$

Combining (9) and (10) we have

$$\lambda_D(E_1, E_2) = \lambda_D(E_1', E_2'). \tag{11}$$

i. e. for purposes of finding the extremal distance we may replace E_1 and E_2 by their boundaries.

3. Determination of a maximal ρ-function. With certain restrictions on E_1, E_2, and D, necessary for the application of the classical existence theorems on harmonic functions, we will now determine a function $\rho \varepsilon \mathrm{P}$ for which the quantity (3) actually attains a maximum. Suppose $D^* = D \cap C(E_1 \cup E_2)$ is a region whose boundary consists of a finite number of analytic curves. Then it is known that there exists a function $u(x, y)$ harmonic in D^* with the boundary values 1 on E_1', 0 on E_2' and normal derivative zero on D'. *We assert that*

$$\rho_0(z) = (u_x^2 + u_y^2)^{\frac{1}{2}} = |\operatorname{grad} u| \tag{12}$$

maximizes (3).

To prove this we observe that all but a finite number of the level curves of the conjugate harmonic function v of u must join E_1' and E_2'. In fact, $\operatorname{grad} u$ can vanish only at a finite number of points, and u is monotonic on any curve $v = \text{const.}$ not containing such a point, which implies that the curve cannot be closed nor have both endpoints on E_1' or both on E_2'. It can of course not continue endlessly in the interior of D^*, since the existence of an accumulation point would then yield $v \equiv \text{const.}$ Also, a level curve of v with $\operatorname{grad} u \neq 0$ is disjoint from D', since $\partial u/\partial n$ was assumed to vanish there.

Now let $\rho(z)$ be any member of the class ρ, normalized as in (4′). Integrating along any level curve of v joining E_1' and E_2' we have

$$\int_{v=c} \rho/\rho_0 \, du = \int_{v=c} \rho/(\partial u/\partial s) du = \int_{v=c} \rho \, |dz| \geq 1 = \int_{v=c} du; \tag{13}$$

hence, since the number of exceptional curves is finite,

$$\iint_{D^*} \rho/\rho_0 \, dudv \geq \iint_{D^*} dudv. \tag{14}$$

But from this it follows that

$$\iint_{D^*} (\rho/\rho_0 - 1)^2 dudv \leq \iint_{D^*} \rho^2/\rho_0^2 \, dudv - \iint_{D^*} dudv, \tag{15}$$

i. e. expressed in x and y, $\rho_0{}^2$ being the Jacobian,

$$(16) \qquad 0 \leq \iint_{D^*} (\rho - \rho_0)^2 dxdy \leq A_\rho(D^*) - A_{\rho_0}(D^*).$$

On account of the form (4′) of the definition of extremal distance, and (11), this means that

$$(17) \qquad \lambda_D(E_1, E_2) = \lambda_{D^*}(E_1', E_2')$$
$$= 1/A_{\rho_0}(D^*) = 1/\iint_{D^*} (u_x{}^2 + u_y{}^2)\, dxdy = 1/D(u).$$

Here and below $D(u)$ denotes the Dirichlet integral taken over

$$D^* = D \cap C(E_1 \cup E_2).$$

This completes the proof of (12).

We may observe that in the restricted case considered in this section, and actually by a suitable limit process in a more general case, the relation (17) could be used as the definition of extremal distance. In view of the electrostatic interpretation, λ could then be called the *resistance* between E_1 and E_2.

It is easy to extend the validity of (12) to the case where E_1, E_2, and D are still bounded by a finite number of analytic curves, but D^* is not connected. The set D^* will then be the sum of a finite number n of regions D_i. For each one of those components D_i, say $D_1, \cdots, D_m$, $m \leq n$, whose boundary contains both a part $E_1{}^i$ of E_1 and a part $E_2{}^i$ of E_2, there exists an extremal distance for which we obtain by (4′)

$$(4'') \qquad 1/\lambda_{D_i}(E_1, E_2) = 1/\lambda_{D_i}(E_1{}^i, E_2{}^i) = \inf_{\rho_i} A_{\rho_i}(D_i), \qquad L_{\rho_i}(\gamma_i) \geq 1.$$

The remaining $D_{m+1}, \cdots, D_n$ have no influence upon the extremal distance between E_1 and E_2, since no curve γ will pass through them. Setting $\rho = \rho_i$ for $i = 1, \cdots, m$, and $\rho \equiv 0$ for $i > m$, we have $L_\rho(\gamma) \geq 1$; hence [3]

$$(18) \quad 1/\lambda_D(E_1, E_2) = \inf_\rho A_\rho(D)$$
$$= \sum_{i=1}^m \inf_{\rho_i} A_{\rho_i}(D_i) = \sum_{i=1}^m 1/\lambda_{D_i}(E_1{}^i, E_2{}^i) = \sum_{i=1}^m 1/\lambda_D(E_1{}^i, E_2{}^i),$$

and by (17) in evident notation

$$(17') \qquad 1/\lambda_D(E_1, E_2) = \sum_{i=1}^m D(u_i) = D(u).$$

[3] The last member of equation (18) has been inserted only for future reference.

4. Continuity of λ. If the sets E_1 and E_2 do not have the simple structure assumed in section **3**, we cannot in general find a harmonic function in D^* with preassigned boundary values and hence the extremal ρ-function, if it exists in the general case, cannot be determined as above. To this end we will prove the following lemma, which enables us to extend many of the subsequent results on extremal distance to any compact sets.

LEMMA. *Let D be a region, and E and F two disjoint compact subsets of D. Then if $\{E_n, F_n\}$ is a sequence of compact subsets of D, covering E and F respectively and converging to E and F (i. e. given $\epsilon > 0$ there exists N such that for $n > N$ every point of E_n and F_n is within distance ϵ of some point of E and F, respectively), we have*

$$\lim_{n\to\infty} \lambda_D(E_n, F_n) = \lambda_D(E, F). \tag{20}$$

We consider the normalized definition (4′), and begin by proving[4] that for any given ρ, the condition $L_\rho(\gamma) \geq 1$ for all γ joining E and F implies that, given $\epsilon > 0$, for $n > n_\epsilon$,

$$L_\rho(\gamma') > 1 - \epsilon \tag{21}$$

for all curves, denoted by γ', joining E_n and F_n. Let z_0 be any point of E, let C_r be the circle $|z - z_0| = r$ for $0 < r \leq k$, where k is such that $C_k \subset D$, and $f(r) = \inf_\gamma L_\rho(\gamma)$ for γ joining F and C_r. We wish to prove that

$$a = \lim_{r\to 0} f(r) \geq 1. \tag{22}$$

From this (21) follows, since E is compact, and the argument can be reapplied to F.

Suppose $a < 1$. Since ρ is square integrable on D we see, by an application of Schwarz' inequality, that for any given $d > 0$ the set $S(d)$ of values of r for which

$$\int_{C_r} \rho \, |dz| < d, \tag{23}$$

has $r = 0$ as a point of accumulation. Hence we can select a sequence $\{r_n\}$, $n = 1, 2, \cdots$ decreasing to zero, $r_1 \leq k$, such that

$$\sum_{n=1}^{\infty} \int_{C_{r_n}} \rho \, |dz| < (1 - a)/2.$$

---

[4] I am indebted to Professor Beurling for an unpublished communication containing the argument that follows.

But, for $\bar{\gamma}$ joining $C_{r_{n+1}}$ and C_{r_n},

$$\inf L_\rho(\bar{\gamma}) \leq f(r_{n+1}) - f(r_n)$$

Hence a curve γ can be constructed, joining F and z_0, such that

$$L_\rho(\gamma) \leq f(r_1) + \sum_{n=1}^{\infty} [f(r_{n+1}) - f(r_n)] + \sum_{n=1}^{\infty} \int_{C_{r_n}} \rho \, | dz | + (1-a)/4$$
$$\leq a + (1-a)/2 + (1-a)/4 = (a+3)/4 < 1.$$

This contradiction proves (22).

To obtain (20) from (21) let us first assume that $\lambda(E, F)$ is finite. Given ϵ, $0 < \epsilon < \lambda(E, F)$, by (4′) there is a ρ for which

$$1/A_\rho(D) > \lambda(E, F) - \epsilon.$$

For this ρ, and $n > n_\epsilon$, by (21) there exist E_n, F_n such that $L_\rho(\gamma') > 1 - \epsilon$ for all γ'. Hence the function $\beta = \rho/(1-\epsilon)$ is one satisfying the normalization (4′) in evaluating $\lambda(E_n, F_n)$. We have

$$\lambda(E_n, F_n) \geq 1/A_\beta(D) = (1-\epsilon)^2/A_\rho(D) > (1-\epsilon)^2(\lambda(E, F) - \epsilon).$$

If $\lambda(E, F)$ is infinite, given $M > 2$, there is a ρ with $1/A_\rho(D) > M$, $L_\rho \geq 1$, and the analogous reasoning gives

$$\lambda(E_n, F_n) \geq 1/A_\sigma(D) > (1 - 1/M)^2 M, \text{ where } \sigma = \rho/(1 - 1/M).$$

This completes the proof of (20).

In the definition (4) of extremal distance we assumed the sets E_1 and E_2 to be closed. This facilitates the statements and proofs of certain results, but it should be pointed out that the restriction is unessential. The definition (4) remains meaningful for arbitrary bounded sets E_1, E_2 with disjoint closures, and *the extremal distance thus defined is equal to the extremal distance between the closures* $\bar{E}_1$, $\bar{E}_2$.

The reasoning which led to the extension principle (6) immediately gives us the inequality

$$\lambda(E_1, E_2) \geq \lambda(\bar{E}_1, \bar{E}_2). \tag{24}$$

To prove the opposite inequality we use the normalized definition (4′) and hence wish to show that, if Γ denotes the class of rectifiable curves γ joining E_1 and E_2, and $\bar{\Gamma}$ the class of rectifiable curves $\bar{\gamma}$ joining $\bar{E}_1$ and $\bar{E}_2$, then for each fixed ρ,

$$\inf_{\gamma \varepsilon \Gamma} L_\rho(\gamma) \leq \inf_{\bar{\gamma} \varepsilon \bar{\Gamma}} L_\rho(\bar{\gamma}). \tag{25}$$

Let $\bar{\gamma}$ be any curve in $\bar{\Gamma}$, its endpoints $z \varepsilon \bar{E}_1$ and $\zeta \varepsilon \bar{E}_2$. For each positive integer n, let $z_n \varepsilon E_1$ and $\zeta_n \varepsilon E_2$ be points such that $| z - z_n | < 2^{-n}$ and $| \zeta - \zeta_n | < 2^{-n}$. Denote by γ_n the curve composed of the polygonal line $z_n, z_{n+1}, z_{n+2}, \cdots$, the curve $\bar{\gamma}$, and the polygonal line $\cdots, \zeta_{n+2}, \zeta_{n+1}, \zeta_n$. This γ_n belongs to Γ, and $\lim L_\rho(\gamma_n) = L_\rho(\bar{\gamma})$ as $n \to \infty$, which proves (25); hence

$$\lambda(E_1, E_2) = \lambda(\bar{E}_1, \bar{E}_2). \tag{26}$$

II. Representation of the Extremal Distance in Terms of a Generalized Potential.

We will now derive a representation of the extremal distance,[5] which will in particular be useful for obtaining estimations.

Let the region D and the compact subsets E_1 and E_2 be bounded by a finite number of analytic curves. Let L be a straight line intersecting D, and denote by $\bar{z}$, $\bar{E}_1$ etc. the points and sets symmetric to z, E_1 etc. with respect to L. Consider the set $E = (E_1 \cap \bar{E}_1) \cup ((E_2 \cap \bar{E}_2)$. If we identify symmetric boundary points of E, a finite number of Riemann surfaces are formed. We are going to apply to the set $D \cap C(E_1 \cup E_2)$, which is now contained in the Riemann surfaces, certain methods similar to those of logarithmic potential theory in the plane. The reader may in a first reading wish to follow the argument in the plane case and may do so by assuming E to be empty or to be situated on the line L. For the applications in Chapter III, the general case is needed, however.

For simplicity we will denote any one of the Riemann surfaces constructed above by D, and by E_1, E_2 and E the intersections of the original sets E_1, E_2 and E with this D. By the existence theorem for abelian integrals of the third kind there is a function $G(\zeta, z_1, z_2)$ with the following properties: Given any two distinct points z_1, z_2 of D, the difference

$$G(\zeta, z_1, z_2) - \log (| \zeta - z_2 | / | \zeta - z_1 |) \tag{1}$$

is harmonic for $\zeta \varepsilon D$; when $\zeta \varepsilon E$, $G(\zeta, z_1, z_2) = G(\bar{\zeta}, z_1, z_2)$ and

$$\partial G(\zeta, z_1, z_2)/\partial n = - \partial G(\bar{\zeta}, z_1, z_2)/\partial n;$$

and $\partial G/\partial n = 0$ on the remaining boundary of D. Since G is only determined up to a constant, we normalize it by requiring the difference (1) to vanish

[5] The possibility of such a representation was suggested to me by Professor Ahlfors.

at an auxiliary point $\zeta = z_0$ on $L \cup E$. For later use we observe that the relations

$$(2) \qquad G(\zeta, z_1, z_2) + G(\zeta, z_2, z_1) = 0.$$

$$(3) \qquad G(\zeta, z_1, z_2) + G(\zeta, z_2, z_3) + G(\zeta, z_3, z_1) = 0,$$

hold for $z_3 \varepsilon D$, $z_3 \neq z_1, z_2$. In fact, let $u(\zeta)$ denote for a moment the left member of either (2) or (3); u is harmonic throughout D, $u(z_0) = 0$ and the Dirichlet integral $\int_{D'} u(\partial u/\partial n) \mid d\zeta \mid$ vanishes due to the properties of G on the boundary D' of D. Hence u is identically zero. To see that $G(\zeta, z_1, z_2)$ is harmonic also in z_1 and z_2 we choose a point z_4 in D, distinct from z_1, z_2 and z_3, and conclude from

$$\int_{D'} [G(\zeta, z_1, z_2) \partial G(\zeta, z_3, z_4)/\partial n - G(\zeta, z_3, z_4) \partial G(\zeta, z_1, z_2)/\partial n] \mid d\zeta \mid = 0$$

that

$$(4) \qquad G(z_1, z_3, z_4) - G(z_2, z_3, z_4) - G(z_3, z_1, z_2) + G(z_4, z_1, z_2) = 0$$

Let M_i be the set of all Borel distributions μ_i on the boundary E_i' of E_i with $\mu_i(E_i') = 1$, $i = 1, 2$, i. e. measures for which every open set is measurable. For any $z \varepsilon D$, the abstract Lebesgue integral of $\log \mid \zeta - z \mid$ over E_i' with respect to any such unit distribution is well defined. Thus we may consider the function

$$(5) \qquad p(z_1, z_2) = \int_{E_1'} G(\zeta, z_1, z_2) \, d\mu_1(\zeta) - \int_{E_2'} G(\zeta, z_1, z_2) \, d\mu_2(\zeta)$$

for any $\mu_1 \varepsilon M_1$, $\mu_2 \varepsilon M_2$ and all values of z_1 and z_2 in D except those for which both points fall simultaneously on E_1' or simultaneously on E_2'. Differentiating under the integral signs we find $p(z_1, z_2)$ to be harmonic in both variables in $D \cap C(E_1' \cup E_2')$. When $z_1 \varepsilon E_1'$ or $z_2 \varepsilon E_2'$ or both, $p(z_1, z_2)$ is lower semi-continuous. To see this we may consider the truncated functions

$$p_n(z_1, z_2) = \int_{E_1'} \min[G(\zeta, z_1, z_2), n] \, d\mu_1(\zeta) - \int_{E_2'} \max[G(\zeta, z_1, z_2), -n] \, d\mu_2(\zeta),$$

which are continuous and increase to $p(z_1, z_2)$ (which may be $+\infty$) as $n \to \infty$. Analogously, if $z_1 \varepsilon E_2'$ or $z_2 \varepsilon E_1'$ or both, we find $p(z_1, z_2)$ to be upper semi-continuous. In particular we conclude that, for each fixed pair of unit distributions, the corresponding function $p(z_1, z_2)$ will attain a minimum for $z_1 \varepsilon E_1'$ and $z_2 \varepsilon E_2'$.

The main theorem of this chapter can now be expressed as follows:

THEOREM. *Let the region D, on a Riemann surface, and the disjoint compact subsets E_1 and E_2 of D be bounded by a finite number of analytic curves. For each pair μ_1, μ_2 of Borel unit distributions on the boundaries E_1', E_2' of E_1, E_2, let the function $p(z_1, z_2)$ be defined by* (5), *where the kernel $G(\zeta, z_1, z_2)$ is defined by* (1). *Then there exists among these pairs of distributions a pair for which the quantity*

$$d = d(\mu_1, \mu_2) = \min_{z_i \in E_i'} p(z_1, z_2)$$

is maximal, and this maximum is equal to 2π times the extremal distance between E_1 and E_2:

$$\lambda_D(E_1, E_2) = 1/2\pi \max_{\mu_i \in M_i} d(\mu_1, \mu_2) = 1/2\pi \max_{\mu_i \in M_i} \min_{z_i \in E_i'} p(z_1, z_2). \tag{6}$$

To prove (6) we first wish to show that

$$\lambda(E_1, E_2) \geq 1/2\pi\, d(\mu_1, \mu_2) \tag{7}$$

for all $\mu_i \in M_i$, and then construct the extremal distributions. Given any fixed pair $\mu_i \in M_i$, it is possible to fix z_2 on E_2' so that $p(z, z_2)$ is still defined for $z \in E_2'$ and non-positive there. In fact, if z_3 is any fixed point in $D \cap C(E_2')$ we can, by the upper semicontinuity of p, choose z_2 to be a point of E_2' at which $p(z, z_3)$ attains its maximum; since by (3) and (2)

$$p(z, z_2) = p(z, z_3) - p(z_2, z_3), \tag{8}$$

the desired non-positiveness follows. Given a positive number ϵ smaller than $d/2$, denote by E_1^* the set where $p(z, z_2) \geqq d - \epsilon$, and by E_2^* the set where $p(z, z_2) \leq \epsilon$. The set E_1^* contains E_1' by the definition of d, and E_2^* contains E_2' by our choice of z_2. By the lower semi-continuity of p on E_1' and the upper semi-continuity on E_2', respectively, the boundaries $E_1^{*\prime}$ and $E_2^{*\prime}$ of E_1^* and E_2^* are disjoint from E_1' and E_2'; hence, as level curves of a harmonic function, they are composed of a finite number of analytic curves. The function

$$p^*(z) = (p(z, z_2) - \epsilon)/(d - 2\epsilon), \tag{9}$$

harmonic in $D \cap C(E_1^* \cup E_2^*)$, has boundary values 0 on $E_1^{*\prime}$ and 1 on $E_2^{*\prime}$, and normal derivative 0 on the remaining boundary. By Ch. I, sec. 3, we conclude that

$$\lambda(E_1^*, E_2^*) = 1/D(p^*) = (d - 2\epsilon)^2/D(p). \tag{10}$$

But for any c, $0 < c < d$, we have, n denoting the inner normal,

$$
\begin{aligned}
(11) \quad D(p) &= -d\int_{p=c} \partial p/\partial n \mid dz \mid = -d\int_{p=c} \mid dz \mid \int_{E_1'} \partial G(\zeta_1, z, z_2)/\partial n \, d\mu_1(\zeta_1) \\
&\quad + d\int_{p=c} \mid dz \mid \int_{E_2'} \partial G(\zeta_2, z, z_2)/\partial n \, d\mu_2(\zeta_2) \\
&= -d\int_{E_1'} d\mu_1(\zeta_1) \int_{p=c} \partial G(\zeta_1, z, z_2)/\partial n \mid dz \mid \\
&\quad + d\int_{E_2'} d\mu_2(\zeta_2) \int_{p=c} \partial G(\zeta_2, z, z_2)/\partial n \mid dz \mid \\
&= d\int_{E_1'} d\mu_1(\zeta_1) \int_{E_2'} d\mu_2(\zeta_2) \int_{p=c} [\partial G(\zeta_2, z, z_2)/\partial n - \partial G(\zeta_1, z, z_2)/\partial n] \mid dz \mid \\
&= -d\int_{E_1'} d\mu_1(\zeta_1) \int_{E_2'} d\mu_2(\zeta_2) \int_{p=c} \partial G(z, \zeta_1, \zeta_2)/\partial n \mid dz \mid = 2\pi d,
\end{aligned}
$$

where the change of the order of integration is justified by Fubini's theorem, the next to the last step is a consequence of (4), and the last equality is verified separately for each of the topologically different mutual positions of the sets E_1, E_2 and the points ζ_1, ζ_2 on them. Substituting (11) into (10) we have, on account of the extension principle (6), Chapter I,

$$\lambda(E_1, E_2) \geq \lambda(E_1^*, E_2^*) = (d - 2\epsilon)^2/2\pi d,$$

hence,

$$(7) \qquad \lambda(E_1, E_2) \geq d/2\pi.$$

To find distributions for which equality holds in (7), we consider the function $u(\zeta)$ which is harmonic on the open set $D^* = D \cap C(E_1 \cup E_2)$, takes the boundary values 0 on E_2' and $\lambda(E_1, E_2)$ on E_1', and has normal derivative zero on D'. Denoting by n the inner normal, we define for all subsets e of E_1' and E_2' respectively,

$$
(12) \qquad
\begin{aligned}
\mu_1(e) &= \int_e \mid \partial u/\partial n \mid \mid d\zeta \mid \\
\mu_2(e) &= \int_e \mid \partial u/\partial n \mid \mid d\zeta \mid .
\end{aligned}
$$

By (17), Ch. I, for $i = 1, 2$,

$$
\begin{aligned}
\int_{E_i'} \mid \partial u/\partial n \mid \mid d\zeta \mid &= -1/\lambda \int_{E_1'} u(\partial u/\partial n) \mid d\zeta \mid = D(u)/\lambda \\
&= \lambda^2 D(u/\lambda)/\lambda = 1.
\end{aligned}
$$

We will show that for the corresponding function p the relation

$$(13) \qquad p(z_1, z_2) = -\int_{E_1' \cup E_2'} G(\zeta, z_1, z_2) \partial u(\zeta)/\partial n \mid d\zeta \mid = 2\pi \cdot \lambda(E_1, E_2)$$

holds for all $z_1 \varepsilon E_1'$, $z_2 \varepsilon E_2'$, which in particular together with (7) will imply (6).

First let us assume z_1 and z_2 to be interior to D^*. If C_i are circles in D^* with centers z_i and radii r, we apply Green's formula in each component of D^* and obtain by adding,

$$(14) \qquad 1/2\pi \int_{D' \cup E_1' \cup E_2' \cup C_1 \cup C_2} (G\partial u/\partial n - u\partial G/\partial n) \, | \, d\zeta \, | = 0$$

which for $r \to 0$ takes the form

$$(15) \qquad u(z_1) - u(z_2) = -1/2\pi \int_{E_1' \cup E_2'} G\partial u/\partial n \, |d\zeta| = 1/2\pi \quad p(z_1, z_2).$$

To extend the validity of (15) to the case where z_i are on E_i' and thus complete the proof of (13), we only have to observe that $p(z_1, z_2)$ is continuous in z_1 and z_2 on the closure of D^*. This follows from the fact that $p(z_1, z_2)$ differs by a continuous function from the ordinary logarithmic potentials of the given distributions with continuous densities.[6]

III. Estimations by Transportation and Projection.

1. **The symmetrization theorem for the entire plane.** The original definition (4), Chapter I, includes the simplest lower estimation of $\lambda_D(E_1, E_2)$: If we choose a particular function $\rho(z)$, the value of the quantity (3) will not exceed λ_D.

The extension principle (6), Chapter I, yields both an upper and a lower estimation: We may construct within E_1 and E_2 sets whose extremal distance we are able to compute explicitly, and we may cover E_1 and E_2 by sets with this property.

In this chapter we shall show how the representation (6), Chapter II, enables us to find interesting upper estimations by deforming and moving the sets E_i, e. g. placing them in symmetric positions with respect to a given axis or projecting them upon a given curve. We commence with the following

THEOREM.[7] *Let E_1 and E_2 be two disjoint compact subsets of the z-plane, D. For every $r \geq 0$ denote by C_r the circle $|z| = r$, and by $\alpha_1(r)$ and $\alpha_2(r)$*

---

[6] See O. D. Kellogg, *Foundations of Potential Theory*, 1929, Chapter VI.

[7] A theorem of the same nature as this for another type of circular symmetrization is given, in the case where the sets are bounded by analytic curves, by G. Pólya, *Comptes Rendus*, Paris, 1950, t. 230, p. 25. Also compare G. Pólya and G. Szegö, *American Journal of Mathematics*, vol. 67 (1945), p. 1-32. Their proof is based on an entirely different idea.

the angular Lebesgue measures of the sets $E_1 \cap C_r$ *and* $E_2 \cap C_r$, *respectively. If the sets* $\tilde{E}_1$ *and* $\tilde{E}_2$ *are defined by the inequalities*

$$(1)\qquad \begin{aligned} \tilde{E}_1:&\quad \pi - \tfrac{1}{2}\alpha_1(r) \leq \phi(r) \leq \tfrac{1}{2}\alpha_1(r) + \pi, \\ \tilde{E}_2:&\quad -\tfrac{1}{2}\alpha_2(r) \leq \phi(r) \leq \tfrac{1}{2}\alpha_2(r), \end{aligned}$$

ϕ *and* r *being polar coordinates in* D, *then* [8]

$$(2)\qquad \lambda_D(E_1, E_2) \leq \lambda_D(\tilde{E}_1, \tilde{E}_2).$$

Our proof of this theorem rests essentially upon

LEMMA 1. *Let* E_1 *and* E_2 *be two disjoint compact sets with boundaries composed of a finite number of analytic curves. Let* L *be a directed straight line through the origin; denote the half plane to the right of* L *by* H, *the left by* $\bar{H}$. *If* A *is any set in the plane, denote by* $\bar{A}$ *the set symmetric to* A *with respect to* L. *Define the sets* $E_1{}^*$ *and* $E_2{}^*$ *as follows:* [9]

$$(3)\qquad \begin{aligned} E_1{}^* &= (E_1 \cap \bar{E}_1) \cup (E_1 \cap \bar{H}) \cup (\bar{E}_1 \cap \bar{H}) \\ E_2{}^* &= (E_2 \cap \bar{E}_2) \cup (E_2 \cap H) \cup (\bar{E}_2 \cap H). \end{aligned}$$

Then

$$(4)\qquad \lambda_D(E_1, E_2) \leq \lambda_D(E_1{}^*, E_2{}^*),$$

where D *again is the entire plane.*[10]

We begin by observing that the reasoning leading to the formula (18), Ch. I, remains valid in the case where the D_i are the Riemann surfaces intro-

---

[8] It is not obvious that the sets $\tilde{E}_1$ and $\tilde{E}_2$ are closed. Those who wish, in defining extremal distance, to restrict themselves to closed sets, may prove the closedness as follows:

Consider a sequence $r_n e^{i\phi_n}$ of points of $\tilde{E}_1$, converging to a point $r_0 e^{i\phi_0}$. It follows from the definition of $\tilde{E}_1$ that the corresponding sequence of sets $E_1 \cap C_{r_n}$ has the property that $\liminf \alpha_1(r_n) \geq 2\phi_0$. We wish to prove that this implies $\alpha_1(r_0) \geq 2\phi_0$. Since E_1 is closed, $\alpha_1(r_0)$ is greater than or equal to the measure β of the set of points on $E_1 \cap C_{r_0}$ which are limit points of sequences of points of the sets $E_1 \cap C_{r_n}$. But it is well-known (cf. E. J. McShane, *Integration*, p. 105) that $\beta \geq \limsup \alpha_1(r_n) \geq 2\phi_0$.

A similar remark applies to the theorems in section 2 of this chapter.

Incidentally, the closedness of the original sets E_1 and E_2 in the above theorem was used to assure the measurability of their intersections with the circles C_r.

[9] In words: To obtain $E^*{}_1$ from E_1 we replace by their symmetric images in $\bar{H}$ those parts of $E_1 \cap H$ whose symmetric images do not belong to E_1. Similarly, we move the parts of $E_2 \cap \bar{H}$ whose images do not belong to E_2.

[10] The principle underlying this approach is related to the method of "rearrangement" discussed by Hardy, Littlewood and Pólya in their *Inequalities*, Cambridge, 1934.

duced at the beginning of Ch. II. Hence it is sufficient to prove Lemma 1 for one such Riemann surface. As before, we will denote it by D, and the boundaries of $E_1 \cap D$ and $E_2 \cap D$ by E_1' and E_2', respectively.

Consider an arbitrary fixed pair of unit distributions μ_1, μ_2 on E_1' and E_2'. Define a corresponding pair μ_1^*, μ_2^* on $E_1^{*\prime}$, $E_2^{*\prime}$ as follows:[11]

$$(5)\quad \begin{aligned} &\mu_1^*(A) = \mu_1(A) \text{ for } A \subset (E_1' \cap \bar{E}_1') \cup (E_1' \cap \bar{H}),\\ &\mu_1^*(A) = \mu_1(\bar{A}) \text{ for } A \subset (\bar{E}_1' \cap \bar{H}) \cap C(E_1' \cap \bar{E}_1'),\\ &\mu_2^*(A) = \mu_2(A) \text{ for } A \subset (E_2' \cap \bar{E}_2') \cup (E_2' \cap H),\\ &\mu_2^*(A) = \mu_2(\bar{A}) \text{ for } A \subset (\bar{E}_2' \cap H) \cap C(E_2' \cap \bar{E}_2'). \end{aligned}$$

Similarly, if $z_1 \varepsilon E_1'$, $z_2 \varepsilon E_2'$ is any pair of points on the original sets we define

$$(6)\quad \begin{aligned} &z_1^* = z_1 \text{ for } z_1 \varepsilon (E_1' \cap \bar{E}_1') \cup (E_1' \cap \bar{H}),\\ &z_1^* = \bar{z}_1 \text{ for } z_1 \varepsilon (E_1' \cap H) \cap C(E_1' \cap \bar{E}_1'),\\ &z_2^* = z_2 \text{ for } z_2 \varepsilon (E_2' \cap \bar{E}_2') \cup (E_2' \cap H),\\ &z_2^* = \bar{z}_2 \text{ for } z_2 \varepsilon (E_2' \cap \bar{H}) \cap C(E_2' \cap \bar{E}_2'). \end{aligned}$$

Setting

$$(7)\quad \int_{E_1^{*\prime}} G(\zeta, z_1, z_2)\, d\mu_1^*(\zeta) - \int_{E_2^{*\prime}} G(\zeta, z_1, z_2)\, d\mu_2^*(\zeta) = p^*(z_1, z_2),$$

we wish to show that

$$(8)\quad p(z_1, z_2) \leq p^*(z_1^*, z_2^*).$$

Since (6) sets up a one-to-one correspondence between the pairs of points on E_1', E_2' and $E_1^{*\prime}$, $E_2^{*\prime}$, (8) implies that

$$(9)\quad \min_{z_i \varepsilon E_i'} p(z_1, z_2) \leq \min_{z_i \varepsilon E_i^{*\prime}} p^*(z_1, z_2)$$

for each pair of distributions; hence by (6), Ch. II, and (11), Ch. I,

$$(10)\quad \lambda_D(E_1, E_2) \leq \lambda_D(E_1^*, E_2^*).$$

We shall collect here a few properties of the function G, which we will need in the verification of (8): If z_1 and z_2 are interior points of $D \cap C(E)$ (we recall that $E = (E_1 \cap \bar{E}_1) \cup (E_2 \cap \bar{E}_2)$), the function

$$(11)\quad G(\bar{\zeta}, z_1, z_2) - G(\zeta, z_1, z_2)$$

[11] In words: if the set A moves in the transformation (3), the mass on A moves with it.

is a harmonic function of ζ in D except for positive poles at $\bar{z}_1$ and z_2 and negative poles at z_1 and $\bar{z}_2$, and vanishes on $E \cup L$. Hence it is non-positive if ζ and z_1 are on the same side of L and z_2 on the opposite side, and non-negative if ζ and z_2 are on the same side of L and z_1 on the opposite side. If z_1 is a point of E, the function (11) is singular only at z_2 and $\bar{z}_2$ (since on the Riemann surface the opposite poles at z_1 and $\bar{z}_1$ coincide), hence non-negative as soon as ζ and z_2 are on the same side of L and non-positive if they are on opposite sides; and analogously for $z_2 \in E$. If both $z_1 \in E$ and $z_2 \in E$, the function (11) vanishes identically.

If z_1 is an interior point and z_2 either an interior point or a point of E, the function

$$G(\zeta, \bar{z}_1, z_2) - G(\zeta, z_1, z_2) \tag{12}$$

is, by (3) and (2), Ch. II, equal to $G(\zeta, \bar{z}_1, z_1)$. On $E \cup L$,

$$G(\zeta, \bar{z}_1, z_1) = G(\bar{\zeta}, \bar{z}_1, z_1) = G(\zeta, z_1, \bar{z}_1) = -G(\zeta, \bar{z}_1, z_1)$$

(the middle equality follows from the fact that $G(\bar{\zeta}, \bar{z}_1, z_1) - G(\zeta, z_1, \bar{z}_1)$ is harmonic everywhere and vanishes at z_0), i. e. $G(\zeta, \bar{z}_1, z) = 0$. Hence the function (12) is non-negative when ζ is on the same side of L as $\bar{z}_1$ and non-positive on the opposite side. If $z_1 \in E$, (12) is identically zero. Analogous considerations apply to the function

$$G(\zeta, z_1, \bar{z}_2) - G(\zeta, z_1, z_2). \tag{13}$$

The verification of (8) must be carried out separately for the different possibilities arising from (6). First let us assume that $z_1^* = z_1$, $z_2^* = z_2$; then

$$\begin{aligned} p^*(z_1^*, z_2^*) - p(z_1, z_2) = &\int_{\bar{E}_1' \cap \bar{H} \cap C(E)} G(\zeta, z_1, z_2)\, d\mu_1(\zeta) \\ &- \int_{E_1' \cap H \cap C(E)} G(\zeta, z_1, z_2)\, d\mu_1(\zeta) - \int_{\bar{E}_2' \cap H \cap C(E)} G(\zeta, z_1, z_2)\, d\mu_2(\zeta) \\ &+ \int_{E_2' \cap \bar{H} \cap C(E)} G(\zeta, z_1, z_2)\, d\mu_2(\zeta) \\ = &\int_{E_1' \cap H \cap C(E)} [G(\bar{\zeta}, z_1, z_2) - G(\zeta, z_1, z_2)]\, d\mu_1(\zeta) \\ &- \int_{E_2' \cap \bar{H} \cap C(E)} [G(\bar{\zeta}, z_1, z_2) - G(\zeta, z_1, z_2)]\, d\mu_2(\zeta). \end{aligned} \tag{14}$$

This is seen to be non-negative by the properties of the function (11) discussed above; e. g. in the first integral on the right either $z_1 \in \bar{H} \cap C(E)$,

$z_2 \varepsilon H \cap C(E)$, in which case ζ and z_2 are on the same side and z_1 on the opposite side of L, or one of the points, e. g. z_1, is on E, in which case ζ and z_2 are on the same side of L, or both points are on E. Next suppose $z_1^* = \bar{z}_1$, $z_2^* = z_2$; then

$$(15) \quad p^*(z_1^*, z_2^*) - p(z_1, z_2) = \int_{E_1' \cap H \cap C(E)} [G(\bar{\zeta}, \bar{z}_1, z_2) - G(\zeta, z_1, z_2)] d\mu_1(\zeta)$$

$$+ \int_{(E_1' \cap \bar{H}) \cup (E_1' \cap \bar{E}_1')} [G(\zeta, \bar{z}_1, z_2) - G(\zeta, z_1, z_2)] d\mu_1(\zeta)$$

$$- \int_{(E_2' \cap H) \cup (E_2' \cap \bar{E}_2')} [G(\zeta, \bar{z}_1, z_2) - G(\zeta, z_1, z_2)] d\mu_2(\zeta)$$

$$- \int_{E_2' \cap \bar{H} \cap C(E)} [G(\bar{\zeta}, \bar{z}_1, z_2) - G(\zeta, z_1, z_2)] d\mu_2(\zeta) \geq 0$$

by the properties of (12) and (13), observing the equation

$$G(\bar{\zeta}, \bar{z}_1, z_2) = G(\zeta, z_1, \bar{z}_2).$$

The case $z_1^* = z_1$, $z_2^* = \bar{z}_2$ is analogous to (15). Finally suppose $z_1^* = \bar{z}_1$, $z_2^* = \bar{z}_2$; then

$$(16) \qquad p^*(z_1^*, z_2^*) - p(z_1, z_2) = \int_{(E_1' \cap \bar{H}) \cup (E_1' \cap \bar{E}_1')} [G(\zeta, \bar{z}_1, \bar{z}_2) - G(\zeta, z_1, z_2)] d\mu_1(\zeta)$$

$$- \int_{(E_2' \cap H) \cup (E_2' \cap \bar{E}_2')} [G(\zeta, \bar{z}_1, \bar{z}_2) - G(\zeta, z_1, z_2)] d\mu_2(\zeta) \geq 0$$

by (11), since

$$G(\zeta, \bar{z}_1, \bar{z}_2) = G(\bar{\zeta}, z_1, z_2).$$

The inequality (8) is verified, and lemma 1 proved. We proceed to

LEMMA 2. *Let E_1 and E_2 be two sets of the type defined by the following inequalities:*

$$(17) \qquad \left.\begin{array}{l} (\nu - 1)\delta \leq r \leq \nu\delta \\ \phi^\nu_{i-1} \leq \phi \leq \phi_i^\nu, \quad i = 1, \cdots, m_\nu \end{array}\right\} \nu = 1, \cdots, n,$$

where ϕ_i^ν are numbers in the interval $0 \leq \phi \leq 2\pi$, and $\delta > 0$ (i. e. sets composed by a finite number of "concentric rectangles" with altitude δ). Then E_1 and E_2 can, by a finite sequence of transformations each of which does not decrease the extremal distance, be transformed into their corresponding symmetric images defined by (1).

First we consider, for a fixed integer ν, the subset E_1^ν of E_1 consisting of m_ν rectangles in the annulus $(\nu - 1)\delta \leq r \leq \nu\delta$. We observe that in a

transformation (3) in any line L through the origin the number m_ν of disjoint rectangles remains the same. A point z of $E_1^{\nu *}$ not in E_1^ν belongs to $(\bar{E}_1^\nu \cap \bar{H}) \cap C(E_1^\nu \cap \bar{E}_1^\nu)$; either it is connected with the set $E_1^\nu \cap \bar{E}_1^\nu$, in which case $\bar{z}$ was connected with $E_1^\nu \cap \bar{E}_1^\nu$ before the transformation, or it belongs to a rectangle R not connected with $E_1^\nu \cap \bar{E}_1^\nu$, in which case $\bar{R}$ was a disjoint rectangle of E_1^ν. The same reasoning applies to points of E_1^ν not in $E_1^{\nu *}$. Now if $\phi_1^\nu \leq \phi \leq \phi_2^\nu$ and $\phi^\nu_3 \leq \phi \leq \phi_4^\nu$, with $0 \leq \phi_1^\nu \leq \phi_2^\nu \leq \phi_3^\nu \leq \phi_4^\nu$, are two rectangles of E_1^ν, a transformation (3) in the axis $L: \frac{1}{2}(\phi_1^\nu + \phi_3^\nu)$ is seen to replace them by a rectangle of length $\phi_4^\nu - \phi_3^\nu + \phi_2^\nu - \phi_1^\nu$ and a line segment $\phi = \phi_1^\nu$, $(\nu - 1)\delta \leq r \leq \nu\delta$. By Lemma 1 the extremal distance does not decrease. Now we remove the line segment; by (6), Ch. I, the extremal distance does not decrease. Repeating this process $m_\nu - 1$ times we have transformed E_1^ν into a single rectangle with angular measure equal to that of E_1^ν.

Having done this for all ν in both E_1 and E_2, we wish to transform the sets obtained, say $\hat{E}_1$, $\hat{E}_2$, into their symmetric images (1). Let R_1 be a rectangle of $\hat{E}_1$, the ϕ-coordinate of its center ϕ_1, with $-\pi < \phi_1 < \pi$. Then a transformation (3) in the axis $L: \frac{1}{2}(\phi_1 + \pi)$ will place R_1 symmetrically with respect to the negative real axis. For $R_2 \subset \hat{E}_2$ with center ϕ_2, $0 < \phi_2 < 2\pi$, a transformation (3) in the axis $L: \frac{1}{2}\phi_2$ places R_2 symmetrically with respect to the positive real axis. Using these transformations we can now move the components of $\hat{E}_1$ and $\hat{E}_2$, one at a time, into their symmetric positions. The fact that, by the definition (3), once a rectangle has reached its symmetric position it remains there under any transformation (3) in an axis between 0 and π—and only such axes L are used in this paragraph—completes the proof of Lemma 2.

The symmetrization theorem (2) can now be proved at once: Let the two disjoint compact sets E_1 and E_2 and positive numbers ϵ and M be given. Let the integer R be the radius of a circle with center at the origin, containing E_1 and E_2. For each integer n we consider the set of closed rectangles formed by the circles and rays

$$
\begin{aligned}
& r = \nu/2^n, && \nu = 0, 1, \cdots, 2^n R, \\
(18) \qquad & \\
& \phi = \mu/2^n R, && \mu = 0, 1, \cdots, [2^{n+1}\pi R].
\end{aligned}
$$

Denote by E_1^n and E_2^n the unions of those rectangles which contain points of E_1 and E_2, respectively. For n large enough, E_1^n and E_2^n will be disjoint, and by (20), Ch. I, n can be chosen so large that

$$
(19) \qquad \lambda(E_1^n, E_2^n) > \lambda(E_1, E_2) - \epsilon,
$$

if $\lambda(E_1, E_2)$ is finite, and

$$\lambda(E_1{}^n, E_2{}^n) > M \tag{19'}$$

if $\lambda(E_1, E_2)$ is infinite. The symmetric images $\bar{E}_1{}^n$, $\bar{E}_2{}^n$ (see (1)) of $E_1{}^n$ and $E_2{}^n$ contain $\bar{E}_1$ and $\bar{E}_2$, respectively. Hence, by (6), Ch. I, and Lemma 2,

$$\lambda(\bar{E}_1, \bar{E}_2) \geq \lambda(\bar{E}_1{}^n, \bar{E}_2{}^n) \geq \lambda(E_1{}^n, E_2{}^n) > \lambda(E_1, E_2) - \epsilon, \tag{20}$$

in the finite case, or

$$\lambda(\bar{E}_1, \bar{E}_2) > M \tag{20'}$$

in the infinite case, i. e.

$$\lambda(\bar{E}_1, \bar{E}_2) \geq \lambda(E_1, E_2), \qquad \text{q. e. d.} \tag{2}$$

The extension principle (6), Ch. I, combined with (2) yields the following

COROLLARY. *Let E_1'' be the radial projection of E_1 upon the negative real axis: $-r \varepsilon E_1''$ if $C_r \cap E_1$ is not empty, and E_2'' the corresponding projection of E_2 upon the positive real axis: $r \varepsilon E_2''$ if $C_r \cap E_2$ is not empty. Then*

$$\lambda_D(E_1'', E_2'') \geq \lambda_D(\bar{E}_1, \bar{E}_2) \geq \lambda_D(E_1, E_2), \tag{21}$$

where D is the entire plane.

2. Similar results for other regions. To obtain results similar to the symmetrization theorem in cases where D is not the entire plane, we may utilize a simple reflexion principle exemplified by the proofs of the following theorems.

THEOREM. *Let E_1 and E_2 be two closed sets in the unit circle D, disjoint from each other and from the origin. If $\bar{E}_1$ and $\bar{E}_2$ denote the corresponding sets defined by* (1), *then*

$$\lambda_D(E_1, E_2) \leq \lambda_D(\bar{E}_1, \bar{E}_2). \tag{22}$$

For the proof, let us first assume E_1 and E_2 to be bounded by a finite number of analytic curves. Let $\hat{E}_1$ and $\hat{E}_2$ be the images of E_1 and E_2 in a reflexion in the boundary C of D. Let $u(z)$ be the harmonic function in $D \cap C(E_1 \cup E_2)$ taking the value 1 on E_1, 0 on E_2, and with normal derivative 0 on C. By (17), Ch. I, we have

$$\lambda_D(E_1, E_2) = 1/D_D(u), \tag{23}$$

where $D_D(u)$ denotes the Dirichlet integral of u over $D \cap C(E_1 \cup E_2)$. In

the reflexion, $u(z)$ is extended into a function $u(z)$ with boundary values 1 on $E_1 \cup \hat{E}_1$, 0 on $E_2 \cup \hat{E}_2$, and harmonic in the entire exterior of these sets. Denoting the entire plane by D^*, we then have

$$(24) \quad \lambda_{D^*}(E_1 \cup \hat{E}_1, E_2 \cup \hat{E}_2) = 1/D_{D^*}(u) = 1/2D_D(u) = \tfrac{1}{2}\lambda_D(E_1, E_2).$$

Analogously,

$$(25) \qquad \lambda_{D^*}(\hat{E}_1 \cup \tilde{E}_1, \hat{E}_2 \cup \tilde{E}_2) = \tfrac{1}{2}\lambda_D(\hat{E}_1, \hat{E}_2).$$

But the symmetrization theorem in D^* applies to the left members of these equations; hence (22) follows for the quantities at right.

In the case where E_1 and E_2 are any disjoint closed sets not containing the origin, we cover them by rectangular sets and reason as in (20).

THEOREM. *Let E_1 and E_2 be two disjoint compact subsets of the set D:*

$$(26) \qquad 0 < \phi < \phi_0 < 2\pi, \qquad r > 0, \qquad z = re^{i\phi}.$$

For all $r > 0$ let C_r denote the circle $|z| = r$, and $\alpha_1(r)$ and $\alpha_2(r)$ the angular Lebesgue measures of the sets $E_1 \cap C_r$ and $E_2 \cap C_r$, respectively. If the sets E_1^* and E_2^* are defined by the inequalities

$$(27) \qquad \begin{aligned} E_1^*&:\quad 0 \leq \phi(r) \leq \alpha_1(r) \\ E_2^*&:\quad \phi_0 - \alpha_2(r) \leq \phi(r) \leq \phi_0, \end{aligned}$$

then

$$(28) \qquad \lambda_D(E_1, E_2) \leq \lambda_D(E_1^*, E_2^*).$$

Let us first assume E_1 and E_2 to be bounded by a finite number of analytic curves. If $\phi_0 \neq \pi$, we begin the proof by mapping D upon the upper half plane by the function $z'' = z^{\pi/\phi_0}$. By (5), Ch. I, the sets E_1, E_2, E_1^*, E_2^* go into sets $E_1'', E_2'', E_1^{*\prime\prime}, E_2^{*\prime\prime}$ with the same mutual extremal distances. Now a reflexion in the real axis yields sets $E_1'' \cup \hat{E}_1''$, etc. to which the symmetrization theorem for the plane applies. The relations (24) and (25) are again valid; we obtain our theorem for the half plane, and by the inverse mapping $z = (z'')^{\phi_0/\pi}$ for our original region D. The extension to arbitrary compact sets is performed as before.

The case where D is a half circle or more generally a circular sector is reduced to the above by a reflexion in the circular part of the boundary.

Extensions of the symmetrization theorem along another line are obtained simply by mapping one of the regions D of the above theorems conformally upon the region for which a symmetrization theorem is desired. Then, however, the symmetrization will not in general remain radial and euclidean.

3. Logarithmic capacity. Let E be a compact set bounded by a finite number of regular curves. Green's function $g(z)$ for the complement of E is uniquely determined by the following requirements: It is continuous for all z, harmonic in $C(E)$, zero on E, and

$$g(z) = \log |z| + c + \epsilon(|z|), \tag{35}$$

where $\epsilon(|z|) \to 0$ as $|z| \to \infty$. The *logarithmic capacity* (transfinite diameter, outer radius) of E is

$$\operatorname{Cap}(E) = e^{-c}. \tag{36}$$

In this section we will derive and apply a relation between logarithmic capacity and a modified form of extremal distance.

Let z_0 be a fixed point in the plane, and C_R a circle with center z_0 and radius R, containing E in its interior. Denoting by $\lambda(C_R, E)$ the extremal distance between C_R and E with respect to the interior of C_R, we define

$$\lambda^*(E) = \lim_{R\to\infty} \{\lambda(C_R, E) - 1/2\pi \log R\}. \tag{37}$$

We need not stop here to verify that this limit exists and is finite and independent of z_0, since this will appear from the argument that follows.

Consider the function [12]

$$\rho_0(z) = |\operatorname{grad} g(z)|. \tag{38}$$

For any curve γ joining E and C_R we have

$$\int_\gamma \rho_0 \, |dz| \geq \int_\gamma \partial g/\partial s \, |dz| = \log R + c + \epsilon(R) \tag{39}$$

(independently of z_0, since $\log |z| = \log |z - z_0| + \epsilon(|z - z_0|)$). Further, if $D = (|z - z_0| < R) \cap C(E)$,

$$\begin{aligned} \iint_D \rho_0^2 \, dxdy &= \int_{C_R} g \, \partial g/\partial n \, |dz| \\ &= \int_0^{2\pi} \{\log R + c + \epsilon(R)\}\{1/R + \epsilon(R)\} R \, d\theta \\ &= 2\pi\{\log R + c + \epsilon(R)\}, \end{aligned} \tag{40}$$

again independently of z_0. Hence by the definition of extremal distance, (4) Ch. I,

$$\lambda(C_R, E) \geq 1/2\pi(\log R + c + \epsilon(R)). \tag{41}$$

---

[12] The following reasoning is similar to that of Chap. I, 3, and we may thus omit details. It follows closely the reasoning for the case of a definition analogous to (37) of a "reduced extremal distance" between E and a finite point, appearing in unpublished work by Ahlfors and Beurling.

Now let ρ be any member of the class P, normalized by $L_\rho(\gamma) \geq 1$. If h denotes the conjugate function of g, we have

$$(42) \qquad \int_{h=c} \rho/\rho_0 \, dg = \int_{h=c} \rho \mid dz \mid \geq 1;$$

hence (compare the discussion of the level curves in Ch. I, **3**)

$$(43) \qquad \int\int_D \rho/\rho_0 \, dg dh \geq 2\pi.$$

Using this and (40) we find (since $\rho_0{}^2$ is the Jacobian)

$$(44) \qquad 0 \leq \int\int_D (\rho/\rho_0 - 1/(\log R + c))^2 dg dh$$
$$\leq \int\int_D \rho^2 \, dx dy - 2\pi/(\log R + c) + \epsilon(R),$$

i. e.

$$(45) \qquad \lambda(C_R, E) \leq 1/2\pi \ (\log R + c) + \epsilon(R).$$

Combining (41), (45) and (37) we conclude that

$$(46) \qquad \lambda^*(E) = c/2\pi.$$

We now claim that $\lambda^*(E)$ does not decrease if E is circularly symmetrized (see (1)) in any line L and with any point z_0 as center. In view of the definition (36) this can be expressed as follows:

THEOREM.[18] *The logarithmic capacity of a compact set bounded by a finite number of regular curves does not increase under circular symmetrization.*

We shall not take the time to remove the restriction on the boundary of E, but it should be observed that this restriction is used only to simplify the derivation of (46).

For the proof we choose in the theorem (1) E as E_1 and the circle C_R with center at the given point z_0 as E_2, and observe that in this case the extremal distance with respect to the entire plane is the same as that with respect to the interior of E_2. For each R, $\lambda(C_R, E)$ does not decrease in symmetrization, and C_R is its own symmetric image. By the definition (37), our theorem follows.

UNIVERSITY OF KANSAS.

---

[18] No published proof of this result is known to the author. By comparing the paper by Pólya and Szegö with the indications in the note of Pólya, both referred to in the first footnote of this chapter, one concludes, however, that the result should be familiar to them. Compare also pp. 182-216 in G. Pólya and G. Szegö, "Isoperimetric inequalities in mathematical physics," *Annals of Mathematics Studies*, no. 27, Princeton, 1951, where several results closely connected to the theorems in this chapter are found.

ON GEODESIC TORSIONS AND PARABOLIC AND ASYMPTOTIC CURVES.*

By Philip Hartman and Aurel Wintner.

1. *Geodesic torsion.* In the differential geometry of surfaces, the assumption that a surface has a parametrization of class C^2 seems to be a natural one. In fact, it is this assumption that permits the definition of both of the fundamental forms and of the standard curves on the surface (geodesics, asymptotic line, lines of curvature). Assumptions of a higher degree of differentiability for the surface usually have therefore no geometrical significance. Cf. [6], [10]. In the light of this remark, various questions centering about the notion of "geodesic torsion" will be considered in what follows.

Let $X=(x,y,z)$ be a vector in a 3-dimensional Euclidean space and let, in a sufficiently small open domain in a (u^1,u^2)-plane,

$$S:\quad X=X(u^1,u^2) \tag{1}$$

denote a (portion of a) surface of class C^2. By this is meant that $X(u^1,u^2)$ is a function of class C^2 and that the vector product (X_1,X_2), where $X_i=\partial X/\partial u^i$, does not vanish. The unit vector

$$N=N(u^1,u^2)=(X_1,X_2)/|(X_1,X_2)| \tag{2}$$

is the normal vector on (1). The first and second fundamental forms of (1) are defined by

$$ds^2=dX\cdot dX=g_{ik}du^idu^k \quad\text{and}\quad -dX\cdot dN=h_{ik}du^idu^k, \tag{3}$$

respectively, that is, by

$$g_{ik}=X_i\cdot X_k \text{ and } h_{ik}=-X_i\cdot N_k=X_{ik}\cdot N, \tag{4}$$

where the dots denote scalar multiplications.

Corresponding to any point (u^1,u^2) of S, let $(u^{1\prime},u^{2\prime})$ represent any pair of numbers for which the vector X' defined by $X_i(u^1,u^2)u^{i\prime}$ is of unit length. Define the vector N' by $N_i(u^1,u^2)u^{i\prime}$, the scalar γ by

$$\gamma=\det(X',N,N'), \qquad (|X'|=1), \tag{5}$$

and call $\gamma=\gamma(u^1,u^2;u^{1\prime},u^{2\prime})$ *geodesic torsion.* If a curve Γ: $X=X(s)$

* Received October 25, 1951.

of class C^1 on the surface S passes through the point X of S and has, at that point, the unit tangent vector $X' = dX(s)/ds$, then (5) will be called the geodesic torsion of Γ at the point X.

The geodesic torsion of the point X and the unit vector X' is often defined to be the torsion (at the point X) of a geodesic through X in the direction X'. Two objections can be raised to the latter definition. First, a geodesic on a surface of class C^2 is only of class C^2, while the standard definition of torsion is applicable only to arcs of class C^3. The second objection is that, even if the geodesic is very smooth, its torsion at a point is undefined if the curvature of the geodesic curve vanishes at that point.

The first criticism can be overcome if the word *torsion* of a curve (which need not be on a surface) is defined as in [6], pp. 770-772, where it was shown that torsion can be defined for certain arcs of class C^2 with non-vanishing curvature. In particular, if such an arc has a principal normal of class C^1, then the torsion can be defined geometrically and so as to satisfy the corresponding Frenet equations. As an application of that definition of torsion, suppose that Γ is a geodesic of (1) and has a non-vanishing curvature (at a point, hence near that point). Then it has a principal normal, namely $\pm N$, and the latter is of class C^1 as a function on the arc length on Γ. Since the binormal of Γ is the vector product $(X', \pm N)$, the Frenet equations, as used in [6] (Theorem VI, p. 772), imply the truth of the following assertion:

(I) *On a surface of class C^2, a geodesic arc of non-vanishing curvature possesses a torsion,* (5).

This disposes of the first of the two objections mentioned above. In contrast, the second of those objections cannot be overcome, except by some *ad hoc* definition of what the torsion of a curve should be when the curvature of the latter vanishes. In fact, even on a surface of class C^∞, those points on a geodesic arc, at which the curvature of the latter vanishes, can form a nowhere dense perfect set. Needless to say, a geodesic (or any curve of class C^2 on a surface of class C^2) can have a vanishing curvature at a point only if it is in an asymptotic direction at that point; so that this is the only case excluded in (I).

Whenever the surface (1) is of class C^2, Weingarten's derivation formulae, $N_i = -g^{jk}h_{ij}X_k$, where (g^{ik}) denotes the inverse of the matrix (g_{ik}), are applicable. Hence (5) can be written in the form

$$\gamma = g(g^{2k}h_{ki}u^{i\prime}u^{1\prime} - g^{1k}h_{ki}u^{i\prime}u^{2\prime}), \tag{6}$$

where $g = \det(X_1, X_2, N) = (\det g_{ik})^{\frac{1}{2}} > 0$.

In order to interpret (6), suppose, without loss of generality, the normalizations $g_{11} = g_{22} = 1$, $g_{12} = 0$ and $h_{12} = 0$ at the point (u^1, u^2). Then h_{11} and h_{22} are the principal curvatures, say κ_1 and κ_2, and the directions determined by $(u^{1\prime}, u^{2\prime}) = (1, 0)$ or $0, 1)$ are directions of principal curvature, associated with κ_1 or κ_2, respectively. If $(u^{1\prime}, u^{2\prime}) = (\cos\theta, \sin\theta)$, where θ is the angle from the direction $(1, 0)$ of principal curvature, associated with κ_1, to the direction $(u^{1\prime}, u^{2\prime})$, then (6) reduces to

$$\gamma = (\kappa_2 - \kappa_1)\cos\theta\sin\theta, \tag{7}$$

a formula which is standard (cf. e. g., [3], p. 389) under substantially more severe restrictions than the present assumptions.

It is clear from (7), and from the assumptions under which it was derived above, that the following statement is a corollary:

(II) *If* Γ *is a curve of class* C^1 *on a surface of class* C^2, *then* $\gamma = 0$ *holds at every point of* Γ *if and only if* Γ *is a line of curvature.*

2. *On the Beltrami-Enneper theorem.* A point of (1) is called elliptic, hyperbolic or parabolic according as $\det h_{ik}$ (or, what according to (4) is the same thing, the Gaussian curvature $K = \det h_{ik}/\det g_{ik}$) is positive, negative or 0.

If (u^1, u^2) is a non-elliptic point of (1), then a direction $(u^{1\prime}, u^{2\prime})$ through the point is called asymptotic if

$$h_{ik}u^{i\prime}u^{k\prime} = 0. \tag{8}$$

In the normalization introduced before (7), condition (8) reduces to

$$\kappa_1\cos^2\theta + \kappa_2\sin^2\theta = 0. \tag{9}$$

On the other hand, since the product $\kappa_1\kappa_2$ is the Gaussian curvature, (7) is identical with

$$\gamma^2 = -K \tag{10}$$

if (9) is satisfied. In view of (7), this proves the following assertion:

(III) *If* Γ *is an arc of class* C^1 *on a surface of class* C^2, *then* (10) *holds at every point of* Γ *if and only if every tangent vector of* Γ *either is, or is orthogonal to, an asymptotic direction.*

The theorem of Beltrami-Enneper states that (under certain conditions of smoothness which usually are not, or are erroneously, specified; cf. [6], pp. 773) the relation (10) holds as an identity along asymptotic curves of

non-vanishing curvature. The first assertion of (III) contains this theorem, and (III) avoids the difficulty involved in a statement about the ***torsion*** (not the *geodesic torsion*) of an asymptotic curve; cf. (IV) below. This difficulty arises when the asymptotic curve is only of class C^1 (or if, when it is smoother, it possesses points at which its curvature vanishes). In this regard, cf. [6], pp. 773-774.

3. *On a result of P. Franklin.* A curve Γ on a surface of class C^2 will be called a parabolic curve if every point of Γ is a parabolic point, that is, if $K = 0$ on Γ.

According to P. Franklin [4], pp. 254-256, a "regular" parabolic curve must be a line of curvature. His definition of a "regular" curve Γ on a surface (1) consists of the following specifications: a) no point of Γ is singular, and b) at no point of Γ does a normal section of the surface S have a flex point; finally c) no point of Γ is a flat point of S (that is, a point at which both factors κ_i of $K = \kappa_1\kappa_2$ vanish).

Franklin's paper was reviewed by Cohn-Vossen and by Rinow. The former [2] points out that Frankin's *proof* is not valid, since it contains a formal error, while the latter [9], without commenting on Franklin's proof, remarks that Franklin's final *assertion* (see above) is not surprising, since the set of the conditions a), b), c) which define "regularity" is quite severe. In what follows, there will be clarified the actual situation resulting from the nature of both of these criticisms. On the one hand, it will be shown that if condition a) is interpreted to mean that grad K (exists and is continuous on S and) does not vanish on the parabolic curve Γ: $K = 0$, then Frankin's *proof* can be saved. On the other hand, it will be shown that condition b) is so severe that the final *assertion* becomes practically vacuous. In fact, the situation proves to be as follows:

(*) *On a surface S of class C^3, let a parabolic curve Γ: $K = 0$ on S be such that* grad $K \neq 0$ *on Γ, and suppose that Γ satisfies Franklin's condition* b). *Then Γ is of class C^2 and a plane curve; as a matter of fact, a plane curve along which the tangent plane of S does not vary; in addition, Γ is a line of curvature as well as an asymptotic curve.*

Actually, it cannot even be expected that a parabolic curve will "in general" be a line of curvature. In fact, the notion of a parabolic curve is an intrinsic one, depending only on the metric (g_{ik}), while the notion of a line of curvature depends on the embedding of the metric into the 3-dimensional Euclidean space (that is, on (h_{ik}) as well). The severity of Franklin's con-

ditions is shown by the last italicized statement, a statement which implies that the parabolic curve must become an asymptotic curve (*and* a line of curvature). This situation is the more understandable as the asymptotic curves are the characteristics of the partial differential equation of second order on which the problem of embedding depends.

4. *Asymptotic curves and parabolic curves.* In view of (II) and (III), and of the last italicized statement, (*), which remains to be proved, it seems to be worth while to clarify the relationships between the following three assumptions: (α) Γ *is an asymptotic curve*; (β) Γ *is a line of curvature*; (γ) Γ *is a parabolic curve,* where it is always assumed that Γ is a curve of class C^1 on a surface S of class C^2. It will be shown that (i) *conditions* (α) *and* (β) *imply* (γ) *and that* (ii) *conditions* (α) *and* (γ) *imply* (β), *but that* (iii) *conditions* (β) *and* (γ) *do not imply* (α), *while* (iv) *conditions* (α), (β), (γ) *all are satisfied if and only if* Γ *is a plane curve along which the tangent plane to the surface* S *does not vary.*[1]

The assertions (i) and (ii) are essentially those of Cohn-Vossen [1], pp. 274-275, who, however, involves the extraneous notion of an "envelope" as well as the notion of *torsion*, of an asymptotic curve (and therefore, in particular, heavy restrictions of differentiability).

Proof of (i) *and* (ii). Condition (α) implies, by (III), that $\gamma = \pm(-K)^{\frac{1}{2}}$ on Γ. Hence, $\gamma = 0$ if and only if $K = 0$. It follows therefore from (II) that an asymptotic curve is a line of curvature if and only if it is a parabolic curve.

Proof of (iv). Clearly, condition (α) is equivalent to the assumption of the relation $X' \cdot N' = 0$ along Γ, which means that the tangent vector X' is orthogonal to N' along Γ. On the other hand, the differential equations of Rodrigues for lines of curvatures, $N' + \kappa_i X' = 0$, where κ_i is a principal curvature and $i = 1, 2$, show that X' is parallel to N' along Γ if and only if Γ is a line of curvature. Thus X' is orthogonal to, and at the same time parallel

[1] In his *Introduction to Differential Geometry* (Princeton, 1947), L. P. Eisenhart claims that every plane asymptotic curve is a straight line (p. 249, Ex. 10). That this theorem is false, or that a (smooth) curve Γ satisfying the three conditions (α), (β), (γ) need not be a straight line, is shown by the following example: Consider the curve Γ: $(x, x^2, 0)$ on the surface S: (x, y, z) belonging to $z = (y - x^2)^2$. Along the curve $y = x^2$ of this surface, both z_x and z_y vanish identically. Hence the plane $z = 0$ contains Γ and is the tangent plane to S at every point of Γ. It follows therefore from (iv) [and, of course, from an easy explicit calculation as well] that conditions (α), (β), (γ) are satisfied. But Γ is not a straight line.

to, N' if and only if (α) and (β) hold. However, in this case, since $X' \neq 0$, it follows that $N' \equiv 0$ along the curve Γ. This means that the plane tangent to the surface S is constant along the curve.

Proof of (iii). Conditions (β) and (γ) imply that both $\gamma = 0$ and (10) hold along Γ. Hence (III) shows that Γ is either an asymptotic curve or is orthogonal to an asymptotic direction at every point of Γ. Consequently, (iii) will be proved if it is shown that the second case of this alternative can actually occur. But it is easy to verify that it does occur if the surface (1) is chosen, for instance, as follows: $z = x^2 + y^3$, where $(x, y, z) = X$.

First, the partial derivatives, z_x, z_y and z_{xx}, z_{xy}, z_{yy}, of $z = x^2 + y^3$ are $2x$, $3y^2$ and $2, 0, 6y$, respectively. Hence, from (2)-(4), where $u^1 = x$, $u^2 = v$,

$$(11) \qquad g_{11} = 1 + 4x^2, \qquad g_{12} = 6xy^2, \qquad g_{22} = 1 + 9y^4$$

and, if $\pm(1 + 4x^2 + 9y^4)^{\frac{1}{2}}$ is denoted by j,

$$(12) \qquad h_{11} = 2/j, \qquad h_{12} = 0, \qquad h_{22} = 6y/j.$$

Since $K = \det h_{ik}/\det g_{ik}$, it follows that

$$(13) \qquad K = 12y/(1 + 4x^2 + 9y^4).$$

Consider the curve

$$(14) \qquad \Gamma: \quad x = x, \qquad y = 0, \qquad z = x^2.$$

This curve is on the surface $S: z = x^3 + y^3$. On the other hand, (12) shows that (8), where $(u^1, u^2) = (x, y)$, reduces to $2dx^2 + 6ydy^2 = 0$, a differential equation for $y = y(x)$ which is not satisfied along the curve (14). Hence (14) is not an asymptotic curve. But it is a parabolic curve, since (13) and (14) imply that $K = 0$. Thus all that remains to be ascertained is that (14) is a line of curvature. This follows, however, by observing that, according to (11) and (12), both g_{12} and h_{12} vanish identically along the coordinate axes, $x = 0$ and $y = 0$, which means that the latter are lines of curvature. Hence the assertion follows from the second of the equations (14).

Condition grad $K \neq 0$ of (*) is satisfied in this example, since the partial derivative of (13) with respect to y is $12/(1 + 4x^2)$, hence distinct from 0, along the curve (14).

5. *Proof of* (*). Consider the surface (1) in a Cartesian parametric form $S: z = (x, y)$, where $z(x, y)$ is a function of class C^3 in a vicinity of $(x, y) = (0, 0)$. It can be supposed that the coordinate axes have been

chosen so that $(x,y)=(0,0)$ corresponds to a given point of Γ, and that the unit normal vector at this point is directed along the z-axis. Thus

(15) $$z(0,0)=0 \text{ and } z_x(0,0)=0, \quad z_y(0,0)=0.$$

Suppose that the Gaussian curvature, $K=K(x,y)$, vanishes at $(0,0)$. Then, since

(16) $$K=(z_{xx}z_{yy}-z_{xy}^2)/(1+z_x^2+z_y^2)^2$$

for every (x,y), and since $K(0,0)=0$, it can be supposed, after a rotation of the (x,y)-plane about the origin, that

(17) $$z_{xy}(0,0)=0 \text{ and } z_{yy}(0,0)=0.$$

The last three formula lines show that

(18) $$K_x(0,0)=z_{xx}(0,0)z_{yyx}(0,0) \text{ and } K_y(0,0)=z_{xx}(0,0)z_{yyy}(0,0).$$

Hence, $|\operatorname{grad} K|^2=K_x^2+K_y^2$ will not vanish at $(0,0)$, and therefore at any (x,y) in a vicinity of $(0,0)$, if and only if

(19) $$z_{xx}(0,0)\neq 0$$

holds and not both $z_{yyx}(0,0)$ and $z_{yyy}(0,0)$ vanish. But if assumption b) of (*) is satisfied, then $z_{yyx}(0,0)$ cannot vanish, since

(20) $$z_{yyy}(0,0)=0.$$

In fact, if (20) did not hold, then, since (15) and (17) imply that $z(0,y)=z_{yyy}(0,0)y^3/6+o(|y|^3)$, it would follow that the normal section $x=0$ of S: $z=z(x,y)$ has a flex point at $y=0$.

The equation $K(x,y)=0$, defining a parabolic curve Γ, and the condition that not both K_x and K_y vanish at a point, say $(0,0)$, of Γ show that Γ is a curve of class C^1 and satisfies the non-singular differential equation $K_x dx+K_y dy=0$. According to (18), (19) and (20), this differential equation reduces to $dx/dy=0$ at $(x,y)=(0,0)$. But (17) and (19) imply that $dx/dy=0$ defines the (unique) asymptotic direction at $(0,0)$. Hence, dx/dy is an asymptotic direction at $(0,0)$. Since $(0,0)$ represents an arbitrary point of Γ, it follows that Γ is an asymptotic curve. Hence Γ is of class C^2 and (5) follows from (i), (ii) and (iii).

6. *On a theorem of van Kampen.* Using a fallacious generalization of the Beltrami-Enneper formula, $\tau^2=K$, for the torsion, τ, of an asymptotic curve, van Kampen [8] has arrived at the following result:

(†) *Let S be a surface of class C^2, and P a hyperbolic point $(K<0)$ of S. Let J denote one branch of the intersection of S and of the plane*

tangent to S at P, and suppose that at every point distinct from P the plane curve J has a non-vanishing curvature. Then there exists on S at least one asymptotic curve, say Γ, which passes through P in such a way that J is tangent to Γ at P and lies between Γ and the common tangent; that is, J lies between Γ and the normal section of S determined by the common direction of J and Γ at P.

That van Kampen's generalization of the Beltrami-Enneper formula is false, is seen by comparing it (Theorem (I) in [8]) with Bonnet's formula (cf. (44) below), which connects the curvature and the torsion of a curve drawn through a point of a surface in an asymptotic direction. The error is made when, in the last sentence of the second paragraph of [8], p. 992, van Kampen assumes that he can differentiate a relation (*along* a curve), whereas the relation in question is valid only at *one* point of that curve.

It turns out, however, that van Kampen's final result, represented by the last italicized statement, (†), happens to be correct. Tis will be proved in what follows.

Remark. It will remain undecided whether, in the above wording of (†), the passage "*at least one* asymptotic curve" can be replaced by "*all* asymptotic curves" or, for that matter, by "*the* asymptotic curve." In this regard, cf. the example, given in [7], pp. 153-156, of a surface S of class C^2 having a negative curvature K and containing a point P which issues more than one asymptotic curve (hence a continuum of asymptotic curves) in the same asymptotic direction. In that particular example, there exist asymptotic curves on different sides of their (common) tangent at P but, in contrast to the assumption in (†) above, the curvature of the curve of intersection, J, vanishes at a sequence of points which cluster at P. Hence it remains a question whether or not the non-vanishing of the curvature of J (at those points of J which are distinct from P) implies that all asymptotic curves (touching J at P) are separated by J from the tangent of J at P or, for that matter, that the asymptotic curve (through P and in the direction of J) must be unique.

It may be mentioned that (†) could also be deduced from Beltrami's formula (cf. (50) below), *if* S is of class C^3 and J has a non-vanishing curvature.

7. *Proof of* (†). It can be assumed that S is given in the form $z=z(x,y)$, where $z(x,y)$ is a function of class C^2; that P is at $(x,y)=(0,0)$ and that (15) is satisfied; finally that, since the value of (16) at P is

supposed to be negative, $z_{xx}(0,0) = z_{yy}(0,0) = 0$, while $z_{xy}(0,0)$ is positive. If (by the choice of the units of length along coordinate axes) this positive number is chosen to be 1, it follows that

$$z = z(x, y) = xy + o(x^2 + y^2), \tag{21}$$

where the o-term represents a function of class C^2 in (x, y). The form on the left of (8) reduces at P to $x'y'$, a form which represents positive values for directions corresponding to the first and third quadrants, $(x' > 0, y' > 0)$ and $(x' < 0, y' < 0)$, and negative values for the second and fourth quadrants. In particular, the asymptotic directions through P are the directions of the coordinate axes.

Consider that branch, J, of the intersection of S and of the tangent plane at P (i.e., of the plane $z = 0$) which is tangent to the x-axis. Then, by Lemma (§) of Section 8 below, J is a curve which (for small $|x|$) can be represented in the form

$$J\colon\ y = y(x), \tag{22}$$

where $y(x)$ is a function of class C^1 for all x, and of class C^2 for all $x \neq 0$; cf. the assumption of (†) concerning the non-vanishing of the curvature of J at points distinct from P.

According to (21), and since (22) is a curve passing through $P = (0, 0)$, the partial derivative $z_y(x, y)$ for small positive x is $x + o(x)$ along the curve J. Hence, if x in (22) is chosen to be positive, then $z_y(x, y(x))$ is positive (for small $x > 0$). On the other hand, since the curve (22) is on the surface $z = z(x, y)$ and on the plane $z = 0$,

$$z(x, y(x)) = 0, \tag{23}$$

and differentiation of (23) with respect to x gives

$$z_x(x, y(x)) + z_y(x, y(x))y'(x) = 0 \tag{24}$$

(even if $x = 0$), whence one more differentiation leads (if $x \neq 0$) to

$$z_{xx} + 2z_{xy}y' + z_{yy}y'^2 + z_y y'' = 0, \qquad \text{where } y = y(x). \tag{25}$$

Since the coefficient, z_y, of y'' in (25) was just seen to be positive (for small positive x), it follows that the sum of the first three terms of (25) is positive or negative according as $y''(x)$ is negative or positive for small positive x (in fact, $y''(x) = 0$ is excluded by the assumption of a non-vanishing curvature at the points of (22) distinct from its point belonging to $x = 0$).

Suppose, for instance, that $y''(x) > 0$ (for small positive x). Then $[y(x)] < 0$, if $[y(x)]$ denotes the sum of the first three terms of (25).

On the other hand, (8) shows that if $y = y^*(x)$, $z = z(x, y^*(x))$ is an asymptotic curve, then $[y^*(x)] = 0$. It follows therefore from the signature rule of the four quadrants (described, after (21), with regard to the second fundamental form of the surface at the point P), that, for small positive x, the slope of the relevant asymptotic direction at the point $(x, y(x))$ of the surface is greater than the slope of the curve (21) at the corresponding x. In fact, this is clear for reasons of continuity, since the asymptotic directions at the point $(x, y(x))$ of the surface are close to those at the point $P = (0, 0)$, while the latter directions are those of the coordinate axes, $x = 0$ and $y = 0$.

Accordingly, if α is any point of an interval $0 < \alpha < a$, where $a > 0$ is sufficiently small, then there is an asymptotic curve (at least one), say

$$(26) \qquad \Gamma_\alpha \colon \ y = y(x; \alpha),$$

passing through the point $(\alpha, y(\alpha))$ of the curve (22) in a direction nearly tangent to the direction of the curve (22) at $x = \alpha$, and all functions (26) of x will exist for $0 \leqq x \leqq a$ (if $0 < \alpha < a$, and if a is sufficiently small), finally all these curves (26) will satisfy, with reference to the curve (22), the inequality

$$(27) \qquad y(x; \alpha) > y(x) \ \text{ if } \ \alpha < x \leqq a.$$

Since the surface S is hyperbolic at (hence near) the point $P = (0, 0)$, its asymptotic curves (26) will satisfy, for $0 \leqq x \leqq a$, a differential equation (8), which is of the form

$$(28) \qquad y' = f(x, y),$$

where $f(x, y)$ is continuous (for small $x^2 + y^2$). Hence the standard arguments, which deal with (28) on the basis of equicontinuous functions, show that there exists a sequence of positive numbers $\alpha_1, \alpha_2, \cdots$ satisfying $a > \alpha_n \to 0$, as $n \to \infty$, and having the property that $y(x; \alpha_n)$, the function (27) belonging to $\alpha = \alpha_n$, tends to a limit function, uniformly for $0 \leqq x \leqq a$, as $n \to \infty$. Furthermore, this limit function $y = y^*(x)$ is (as is each of the curves (26)) the graph of the projection on the (x, y)-plane of an asymptotic curve. Finally, it follows from (27) that

$$(29) \qquad y^*(x) \geqq y(x) \qquad \text{for } 0 \leqq x \leqq a.$$

Clearly, (29) completes the proof of the last italicized statement, (†), if the lemma (§) of Section 8 is granted. In fact, while (29) was deduced for the case in which $y''(x) > 0$ for small $x > 0$, the case in which $y''(x) < 0$ for small $x > 0$ (as well as both cases of a small $-x > 0$) can, of course, be treated in the same way.

8. *On the "true" Dupin diagram.* Let T be a plane through a hyperbolic point P of a surface S of class C^2. If T is not the plane tangent to S at P, and if S is a sufficiently small neighborhood of P, then the set ST, along which T intersects S, consists of a Jordan arc of class C^2. This is an immediate consequence of (the C^2-form of) the classical theorem on implicit functions, the exemption of the tangent plane being equivalent to the non-vanishing of the gradient involved. Correspondingly, if T is the tangent plane, then, since the gradient vanishes at P, no general theorem on implicit functions is applicable to either of the branches of which, in view of Dupin's indicatrix, the set ST can be expected to consist. That the essential (but, perhaps against expectation, not all) aspects of what is indicated by Dupin's approximation is nevertheless true, is the content of the first two assertions of the following lemma (the third assertion of which shows that, because of the vanishing of the gradient, the C^2-character of the "implicit functions" can actually be lost).

(§) *If S is a sufficiently small neighborhood of a hyperbolic point P on a surface of class C^2, and if T denotes the plane tangent to S at P, then*

(i) *the intersection ST consists of two Jordan arcs, say J_1 and J_2, each of which contains P in its interior, and P is the only point common to J_1 and J_2;*

(ii) *both plane curves J_1, J_2 have continuous tangents (also at P) and, except possibly at P, continuous curvatures as well;*

(iii) *the curvature of J_i at P need not exist (and, if it exists at P, it need not be continuous at P)*, unless an assumption going beyond the C^2-assumption (such as the C^3-assumption) is required of S.

Assertion (i) and (at least the first part of) assertion (ii) are closely connected with the results of Hadamard [5] on the invariant curves of a surface transformation near a fixed point of hyperbolic type. It is, however, more convenient to prove (i) and (ii) directly, and in a way which, in contrast to Hadamard's procedure, does not depend on the method of successive approximations.

As at the beginning of Section 7, it can be assumed that S is given in the form $z=z(x,y)$, and that T is the (x,y)-plane and P is its point $(0,0)$, but that (21) is replaced by

$$(30) \qquad S\colon\ z=z(x,y)=\tfrac{1}{2}(y^2-x^2)+o(x^2+y^2),$$

where the o-term represents a function which is of class C^2 in (x,y) (in fact, (30) differs from (21) in a rotation about P). According to (30), the

asymptotic directions of S at P, represented by the asymptotes of Dupin's hyperbola at P, are the bisectors $(x \pm y = 0)$ of the coordinate quadrants.

Since T is the plane $z = 0$, its intersection with S is the set of points (x, y) satisfying

$$(31) \qquad ST\colon\ \tfrac{1}{2}(y^2 - x^2) + o(x^2 + y^2) = 0.$$

It is clear from (31) that, if S, or the (x, y)-domain under consideration, is chosen small enough, then every point of ST is contained in one of the four wedges (issuing from $(0, 0)$ and bisected by the four asymptotic half-lines) which are defined by the inequalities $\frac{1}{2}|x| \leqq |y| \leqq 2|x|$.

Consider the wedge contained in the first quadrant of the (x, y)-plane, that is, the wedge

$$(32) \qquad \tfrac{1}{2}x \leqq y \leqq 2x \qquad \text{(so that } x > 0 \text{ and } y > 0 \text{ unless } x = 0 = y).$$

It will be shown that those points (x, y) of ST which are contained in this wedge form (for sufficiently small $x \geqq 0$) a Jordan arc which is representable in the form $y = y(x)$, where $y(x)$ is a (single-valued) function having a continuous first derivative $y'(x)$; that $y'(0)$ (when interpreted as the derivative at $x = 0$ from the right) is 1; finally that the function $y(x)$ has a continuous second derivative $y''(x)$ if $x > 0$. Since (32) could be replaced by any of the four wedges, this will prove assertions (i) and (ii) of the last italicized assertion.

Proof of (i)-(ii). It is clear from (30) that, if x is positive (and small enough), then $z(x, \frac{1}{2}x)$ and $z(x, 2x)$ are of opposite sign. Hence $z(x, y)$ must vanish at least once, say at the ordinate $y = y(x)$, when $x > 0$ is fixed and y varies from the lower to the upper half-line bordering the wedge (32). On the other hand, since the o-term in (30) is a function of class C^2 in (x, y), it is clear from (30) that $z_y(x, y) = y + o(x^2 + y^2)^{\frac{1}{2}}$. Hence, if $x > 0$ is small enough, $z_y(x, y)$ is positive within the wedge. Consequently, the ordinate $y = y(x)$, mentioned before, is unique. It now follows by standard arguments, occurring in the proof of the classical theorems on implicit functions, that the function $y(x)$ is continuous for $x \geqq 0$ and that it has a continuous second derivative for $x > 0$; cf. (24) and (25). That it has a continuous first derivative at $x = 0$ also, follows from (24). In fact, (30) shows that (24) can be written in the form

$$-x + o(x) + (y(x) + o(x))y'(x) = 0$$

for small positive x, and (31) shows that $y(x)/x = 1 + o(1)$.

Proof of (iii). In a neighborhood of $x = 0$, let $f(x)$ be any function

satisfying the following pair of conditions: (a) there exists a continuous second derivative $f''(x)$ (also at $x=0$); (b) if $x \to \pm 0$, then $f(x)=o(x^2)$. In terms of such an $f(x)$, define S by $z=z(x,y)$, where $z(x,y)=xy+f(x)$. Then S is of class C^2, by (a), and (21) is satisfied, by (b). Furthermore, it is seen that (23), the equation defining the intersection ST, splits into J_1: $x=0$ and J_2: $y=f(x)/x$. Hence, in order to conclude the truth of (iii), it is sufficient to observe that there exist two functions, say $f(x)=g(x)$ and $f(x)=h(x)$, which satisfy both (a) and (b) and have the property that the function defined by $y(x)=f(x)/x$ (if $x\neq 0$, and by $y(0)=0$ if $x=0$) has at $x=0$ no second derivative or a discontinuous second derivative according as $f=g$ or $f=h$.

9. *Geodesic curvature and geodesic torsion.* After the italicized assertion, (†), of Section 6, reference was made to a formula due to Bonnet (cf. [3], pp. 397-399). Inasmuch as this formula, which is (44) below, is usually derived in a somewhat roundabout way and without a specification of the assumptions on which it depends, it will be proved in what follows by a more direct approach, leading to a reasonable minimum of the conditions to be required for its validity.

Let S be a surface of class C^2, and Γ: $X=X(s)$ a curve of class C^2 on Γ, where s denotes the arc length on Γ. Define on Γ three, mutually perpendicular, unit vectors V_1, V_2, V_3 (which form a right-hand orthogonal system), by placing

$$(33) \qquad V_1(s)=X'(s), \qquad V_2(s)=N(s), \qquad V_3(s)=(X'(s),N(s)),$$

where the prime denotes differentiation with respect to s and N is the surface normal, (2), expressed along Γ as a function of s. Clearly, all three functions $V_i(s)$ are of class C^1. Hence the three derived vectors, V_i', exist, are continuous, and are linear combinations, with continuous scalar coefficients, of the three vectors V_k. These coefficients can be calculated as is usual for all "derivation formulae." This leads to the well-known "geodesic Frenet equations,"

$$(34) \qquad V_1'=\alpha V_2-\beta V_3, \qquad V_2'=-\alpha V_1+\gamma V_2, \qquad V_3'=\beta V_1-\gamma V_2,$$

where

$$(35) \qquad \alpha=X''\cdot N, \qquad \beta=\det(X',X'',N), \qquad \gamma=\det(X',N,N').$$

The latter γ is identical with the γ in (5), which in Section 1 was defined to be the *geodesic torsion*. Correspondingly, the second of the relations (35) shows that β is identical with the classical ("embedded") definition of the

geodesic curvature. Finally, it is seen from (4) that the first of the relations (35) can be written in the form

$$(36) \qquad \alpha = -X' \cdot N'; \text{ hence } \alpha = h_{ik} u^{i\prime} u^{k\prime},$$

by (3) (so that α is the *normal curvature*). Thus $\alpha = 0$ is equivalent to (8), which is the definition of an asymptotic direction (if any; that is, if $K \leqq 0$).

If points of Γ at which $|X''|$ may vanish are excluded, then the set (33) (in which the assumption $|X''| > 0$ is not needed) can be paralleled by the set consisting of the unit vectors of the tangent, principal normal, and binormal, of Γ, that is, by the set

$$(37) \qquad U_1 = X', \qquad U_2 = |X''|^{-1} X'', \qquad U_3 = |X''|^{-1}(X', X'').$$

Then the "geodesic" Frenet equations, (34), become replaced by the ordinary Frenet equations,

$$(38) \qquad U_1' = \kappa U_2, \qquad U_2' = -\kappa U_1 + \tau U_3, \qquad U_3' = -\tau U_2,$$

and, correspondingly, the data (35) by the (ordinary) *curvature* and the (ordinary) *torsion,*

$$(39) \qquad \kappa = |X''| > 0 \text{ and } \tau = \det(X', X'', X''')/\kappa^2,$$

provided that Γ: $X = X(s)$ (instead of being, as before, just of class C^2) is of class C^3. But the definitions (37) and the first of the relations (38), where $\kappa = |X''| > 0$, do not require the latter proviso, and imply, in view of (37), the identities

$$(40) \quad U_1 = V_1, \quad U_2 = V_2 \cos\omega + V_3 \sin\omega, \quad U_3 = -V_2 \sin\omega + V_3 \cos\omega,$$

if the (continuous) angular function $\omega = \omega(s)$ is defined (mod 2π) by

$$(41) \qquad \cos\omega = N \cdot U_2, \qquad \sin\omega = -\det(U_1, U_2, N).$$

If this is compared with (37) and (38), it follows that

$$(42) \qquad \alpha = \kappa \cos\omega, \qquad \beta = -\kappa \sin\omega.$$

Since the definitions of α, β and κ imply that $\alpha = 0 = \beta$ if $\kappa = 0$, both equations (42) hold for $\kappa = 0$ also, provided that the angle ω, which the case $\kappa = 0$ of (41) leaves undefined, is considered as arbitrary.

It will now be supposed that $\kappa > 0$, and that Γ is of class C^3. Then (38) is applicable and, in view of (41), the continuous function $\omega = \omega(s)$ is of class C^1, as is $\kappa = \kappa(s)$. Hence, if (40) is differentiated, and if the result is compared with (34) and (38), it follows that

$$(43) \qquad \omega' = \tau - \gamma.$$

A corollary of (36), (42) and (43) is the following pair of facts:

(IV) *Let S be a surface, of class C^3, having no elliptic points (so that $K \leqq 0$ on S), and let Γ be a curve, of class C^2, on S. Then Γ is an asymptotic curve if and only if $\kappa(s) \equiv |\beta(s)|$, where $\kappa(\geqq 0)$ is the curvature, and β the geodesic curvature, of Γ. If, in addition, $\kappa(s) \neq 0$ holds on an asymptotic curve Γ, then $\tau(s) \equiv \gamma(s)$, where $\tau(s)$ is the torsion, and $\gamma(s)$ the geodesic torsion, of Γ.*

Since S is supposed to be of class C^3, the asymptotic curves are of class C^2. Hence, the classical definition, (39), of τ will not in general apply (cf. [6], pp. 773). But if the torsion, τ, is defined as in [6], pp. 770-772, then the asymptotic curve has a torsion at all those points s at which $\kappa(s) \neq 0$. This, and only this, makes meaningful the second assertion of (IV).

Of course, the converse of the second assertion of (IV) is false, that is, the assumption $\tau(s) \equiv \gamma(s)$ along a curve of class C^2 (with non-vanishing curvature) does not imply that the curve is an asymptotic curve. In fact, this identity holds, under the assumption $\kappa(s) > 0$, if and only if $U_3(s) \cdot N(s) =$ const. on Γ (a condition which is satisfied if, for instance, Γ is an asymptotic curve or a geodesic).

Let s be fixed. Then, if $\kappa(s) = |X''(s)|$ vanishes, (35) shows that $\alpha(s) = \beta(s) = 0$. On the other hand, if $\kappa(s) > 0$, then the first of the relations (42) shows that $\alpha(s) \equiv 0$ is equivalent to $\cos\omega(s) \equiv 0$, which implies that $\omega(s)$ is a multiple of π, and that $\omega'(s)$ therefore exists and is 0 (in a vicinity of the fixed s). In other words, $\alpha(s) \equiv 0$ implies that $\kappa \equiv |\beta|$, by the second part of (42), and that $\tau \equiv \gamma$ if $\kappa \not\equiv 0$, by (43); conversely, $\kappa \equiv |\beta|$ implies that $\alpha \equiv 0$. This completes the proof of (IV), since, in view of the remark made after (36), the differential equation of the asymptotic curves is $\alpha = 0$.

10. *On a formula of Bonnet.* It will now be easy to formulate a precise wording of Bonnet's relation (cf. [3], pp. 397-399), referred to at the beginning of Section 9.

(V) *Let S be a surface of class C^3 having no elliptic points (so that $K \leqq 0$), and let Γ be a curve on S which is of class C^3 and has, at some point P, an asymptotic direction and a non-vanishing curvature. Then*

$$(3\tau_0 - \tau)\kappa = \pm 2\tau_0\kappa_0, \qquad (\kappa > 0, \kappa_0 \geqq 0), \tag{44}$$

where κ and τ denote curvature and the torsion, and $\pm\kappa_0$ and $\tau = \pm(-K)^{\frac{1}{2}}$ the geodesic curvature and the geodesic torsion, of Γ at P (so that, if Γ_0 denotes

the asymptotic curve tangent to Γ at P, then κ_0 is the curvature and, if $\kappa_0 > 0$, then τ_0 is the torsion of Γ_0 at P; cf. (IV) above).

When $\kappa\kappa_0 > 0$, the alternative sign $(\pm)$ in (44) depends on Γ; in fact, it will be clear from the proof of (44) that the $+$ or the $-$ holds according as the principal normals of Γ and Γ_0 have common or opposite directions.

If P is a parabolic point of S, then any curve Γ (of class C^3) through P having an asymptotic direction possesses, at P, either a vanishing curvature κ or a vanishing torsion τ, since $\kappa > 0$ and $\tau_0 = \pm(-K)^{\frac{1}{2}} = 0$ imply that $\tau = 0$ in (44).

The proof of the last italicized statement, (V), proceeds as follows:

First, a multiplication of the two determinants in (35) shows that $\beta\gamma$ can be written as the determinant in which the first row is 1, 0, $X'\cdot N'$, the second 0, $X''\cdot N$, $X''\cdot N'$, and the third 0, 1, 0. Hence $\beta\gamma = -X''\cdot N'$. Since differentiation of the first of the relations (36) gives $\alpha' = -X''\cdot N' - X'\cdot N''$, it follows that

$$\alpha' - 2\beta\gamma \tag{45}$$

is identical with $X''\cdot N' - X'\cdot N''$.

Since the curve Γ: $X = X(s) = X(u^1(s), u^2(s))$ is on the surface S: $X = X(u^1, u^2)$, differentiations of (1) and (2) with respect to s show that $Z' = Z_i u^{i\prime}$ and $Z'' = Z_{ik}u^{i\prime}u^{k\prime} + Z_i u^{i\prime\prime}$ hold for $Z = X$ and for $Z = N$ (the subscripts denote partial differentiation with respect to u^i, u^k). Hence, the expression (45) is the sum of

$$(X_{ik}\cdot N_j - X_j\cdot N_{ik})u^{i\prime}u^{k\prime}u^{j\prime} \tag{46}$$

and of the bilinear form $h_{ik}(u^{i\prime\prime}u^{k\prime} - u^{k\prime\prime}u^{i\prime})$, where h_{ik} is the scalar product defined by (3) or (4). This bilinear form vanishes identically, since $h_{ik} = h_{ki}$. Consequently, the function (45) or (46) of s depends only on the point P: $X(s)$ and on the direction, $X'(s)$, of Γ at P.

The identity of the two values (45), (46) was just derived under the hypothesis that Γ is of class C^2. If Γ is of class C^3, so that $\kappa(s)$ and $\omega(s)$ are of class C^1, then (42) shows that (46) can be written in the form

$$\kappa'\cos\omega - (\omega' - 2\gamma)\kappa\sin\omega. \tag{47}$$

Let this be applied to both Γ_0 and Γ, where Γ_0 is an asymptotic curve through a point, P, of S, and Γ is a curve, of class C^3 and of non-vanishing curvature, which is on S and is tangent to Γ_0 at P. In view of the identity of the numbers (45), (46), it is seen that (45) attains, at P, the same

value for Γ as for Γ_0. Hence, if $\kappa, \omega, \gamma, \cdots$ and $\kappa_0, \omega_0, \gamma_0, \cdots$ refer to Γ and Γ_0, respectively, then the value of the expression (45) at the point P is

$$(48) \qquad -2\beta\gamma - 2\kappa_0\tau_0 \sin \omega_0, \qquad (\sin \omega_0 = \pm 1),$$

since $a(s) \equiv 0$, hence $a'(s) \equiv 0$, on Γ_0.

On the other hand, since Γ is of class C^3, another expression for the value of (45) at P is given by (47). In view of (43), and since $\gamma = \tau_0$, the identity of the values (47), (48) means that

$$(49) \qquad \kappa' \cos \omega + (3\tau_0 - \tau)\kappa \sin \omega = 2\kappa_0\tau_0 \sin \omega_0.$$

Finally, since Γ has an asymptotic direction at P, it follows that $a(s) = 0$ at P. Hence, the first relation of (42) and the assumption $\kappa(s) \neq 0$ show that $\cos \omega = 0$, $\sin \omega = \pm 1$. Consequently, (44) follows from (49).

11. *On a formula of Beltrami.* Beltrami's theorem, mentioned at the end of Section 6, states that if Γ is a branch of the intersection, ST, of S and of the plane, T, tangent to S at a hyperbolic point, P, of S, then, in the notations of (V) above,

$$(50) \qquad 3\kappa = 2\kappa_0$$

(cf. [3], p. 398). Formally, this result of Beltrami is a consequence of Bonnet's theorem, since (44) reduces to (50) if $\tau = 0$ (in fact, since $\tau_0 = \pm(-K)^{\frac{1}{2}}$, hence

$$(51) \qquad \tau_0 \neq 0$$

at a hyperbolic point, division by τ_0 is allowed).

Actually, this deduction of (50) from (44) is not legitimate under the assumptions of (V). For, on the one hand, (50) is claimed also for the case $\kappa = 0$, excluded in (38), and, on the other hand, (V) assumes that Γ is of class C^3, whereas, corresponding to (iii) in (§), Section 8, the curve Γ need not be of class C^3 under the C^3-assumption made in (V) for S. It will, however, be shown that the proof of (V) can be adjusted to the case of Beltrami's theorem so as to dispose of the difficulties on both of these accounts; so that (50) holds in its full generality:

(VI) *Let P be a hyperbolic point of a surface S of class C^3, and let Γ denote a branch of the intersection ST, where T is the plane tangent to S at P. Then Γ has at P a (continuous) curvature, $\kappa \geqq 0$, and (50) holds for the curvature, $\kappa_0 \geqq 0$, of the asymptotic curve tangent to Γ at P.*

In order to prove this assertion, (VI), suppose first that $\kappa(s)$, the curvature of Γ, vanishes on a set of points which cluster at P. It then follows from (42) that, for reasons of continuity, $\kappa = \beta = 0$ at P. On the other hand, it was shown in Section 10 that α' exists even under the present assumptions, and so it is seen from (42) that α' vanishes at P. Consequently, the same is true of the expression (45), and therefore of the expression (46) as well, and so of (48). But the vanishing of (48) at P is equivalent to $\kappa_0 = 0$, since (51) is satisfied (cf. the corresponding conclusion in Section 10). Hence (50) is true in the present case.

In the remaining case, $\kappa(s)$ does not vanish near P (it may or may not vanish *at* P), hence (41) defines $\omega(s)$ near P. It follows therefore from an obvious variant of the lemma, (§), of Section 8 that Γ is of class C^3 near P. But $\tau(s) \equiv 0$, since $\Gamma = ST$ is a plane curve. Thus it is clear from (43) that (whether $\kappa(s)$ does or does not vanish at P) it is possible to define ω (*at* P) in such a way that $\omega(s)$ remains of class C^1 after the inclusion of P; in fact,

$$\omega' = -\gamma = -\tau_0 \text{ at } P. \tag{52}$$

On the other hand, since Γ is in the plane tangent to S at P, it is clear that, *near* P, the vector product (U_1, U_2) has the constant direction of the normal to S at P (note that the principal normal, U_2, is undefined *at* P if $\kappa = 0$). It follows therefore from (41) that

$$\sin\omega = \pm 1 \text{ and } \cos\omega = 0 \text{ at } P, \tag{53}$$

and it is clear from (53) that $\alpha(s) = \kappa(s)\cos\omega(s)$ is differentiable at P, having the derivative

$$\alpha' = 0 + \kappa(\cos\omega(s))'_P = -\kappa\omega'\sin\omega \text{ at } P \tag{54}$$

(even though $\kappa(s)$ may not be differentiable at P). From now on, all statements will refer to the point P.

Since (54), (52) and the second of the relations (42) imply that $\alpha' = \kappa\tau_0\sin\omega = -\beta\gamma$, the value of the expression (45) is $3\kappa\tau_0\sin\omega$. On the other hand, the expression (45) is identical with (46) as well as with (48). Consequently, the product on the right of (48) must have the value $3\kappa\tau_0\sin\omega$, just found. In view of (51), this means that $2\kappa_0\sin\omega_0 = 3\kappa\sin\omega$, which, according to (53) and the parenthetical alternative in (48), implies that $|2\kappa_0| = |3\kappa|$. This proves (50), since both curvatures κ, κ_0 are non-negative in (VI).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] S. Cohn-Vossen, "Die parabolische Kurve," *Mathematische Annalen*, vol. 99 (1928), pp. 273-308.

[2] ———, *Zentralblatt für Mathematik*, vol. 9 (1934), pp. 373-374.

[3] G. Darboux, *Leçons sur la théorie générale des surfaces*, vol. 2 (1889).

[4] P. Franklin, "Regions of positive and negative curvature on closed surfaces," *Journal of Mathematics and Physics*, vol. 13 (1934), pp. 253-260.

[5] J. Hadamard, "Sur l'itération et les solutions asymptotiques des équations différentielles," *Bulletin de la Société Mathématique de France*, vol. 29 (1901), pp. 224-228.

[6] P. Hartman and A. Wintner, "On the fundamental equations of differential geometry," *American Journal of Mathematics*, vol. 72 (1950), pp. 757-774.

[7] ——— and A. Wintner, "On the asymptotic curves of a surface," *ibid.*, vol. 73 (1951), pp. 149-172.

[8] E. R. van Kampen, "A remark on asymptotic curves," *ibid.*, vol. 61 (1939), pp. 992-994.

[9] W. Rinow, *Jahrbuch über die Fortschritte der Mathematik*, vol. 60_1 (1934), pp. 613-614.

[10] A. Wintner, "On isometric surfaces," *American Journal of Mathematics*, vol. 74 (1952), pp. 198-214.

ON THE THEORY OF GEODESIC FIELDS.*

By PHILIP HARTMAN and AUREL WINTNER.

1. *Geodesics.* On a sufficiently small (u, v)-domain, say on

$$\mathcal{B}_a: \quad u^2 + v^2 < a^2, \tag{1}$$

consider the Riemannian geometry defined by a line element

$$ds^2 = E(u, v)\,du^2 + 2F(u, v)\,du\,dv + G(u, v)\,dv^2, \tag{2}$$

which is positive definite (i. e.,

$$EG - F^2 > 0 \text{ and } E > 0, \tag{3}$$

hence $G > 0$), but such that the functions E, F, G of (u, v) are just continuous. A geodesic Γ must then be defined as a Jordan arc contained in $\mathcal{B}_a$ and having the property that, if P and Q are the end points of Γ, and if Λ is any Jordan arc joining P to Q within $\mathcal{B}_a$, then

$$\int_\Gamma |\,ds\,| \leqq \int_\Lambda |\,ds\,|.$$

If P and $Q \neq P$ are close enough to the center, $(0, 0)$, of the (u, v)-domain (1), then, according to Hilbert ([9]; cf. [2], pp. 419-438), there exists a geodesic $\Gamma = \Gamma(P, Q)$, which is a rectifiable Jordan arc. (Actually, Hilbert assumes that the line element (2) is embedded into a Euclidean (x, y, z)-space as the $dx^2 + dy^2 + dz^2$ on a surface of class C^1, but this assumption is nowhere used in his proof.)

One of the difficulties is that, no matter how close Q be chosen to a fixed P, the geodesic joining Q to P need not be unique, not even if the coefficient functions of (3), instead of being just continuous, are of class C^1; cf. [7], pp. 132-133. If they are of class C^1, then the Christoffel coefficients $\Gamma^i_{jk} = \Gamma^i_{jk}(u, v)$ exist and are continuous, and, as pointed out in [7], pp. 134-135, all geodesics must satisfy the standard differential equations $u^{i\prime\prime} + \Gamma^i_{jk} u^{j\prime} u^{k\prime} = 0$, where $(u^1, u^2) = (u, v)$.

Since the coefficient functions of (2) will not be required to possess derivatives, no differential equations for the geodesics will be available. Nevertheless, there will be proved several central theorems of the classical theory

* Received November 12, 1951.

(a theory the methods of which assume the coefficient functions of (2) to be of class C^2, at least, and often even smoother). The theorems in question are those of Gauss on orthogonal trajectories, the related theorem of Jacobi concerning multipliers, Riemann's invariant relations which are both necessary and sufficient (Herglotz) for normal geodesic coordinates, and Beltrami's characterization of the non-Euclidean and spherical metrics as geodesic maps of the Euclidean geometry.

Even if the coefficient functions of (2) were assumed to be of class C^1, the metric defined by (2) would not, in general, have a Gaussian curvature $K = K(u, v)$ (in fact, the classical definition of K applies only if E, F, G are of class C^2). In particular, Jacobi's equation of the normal displacements, $d^2n/ds^2 + K(s)n = 0$, which defines conjugate points, is not available.

In problems dealing with the geodesics of a metric (2) in which E, F, G are just continuous, a complication is presented by the circumstance that a geodesic arc Γ cannot be assumed to be a curve of class C^1 or, for that matter, such as to possess a tangent at each of its points (rather than just almost everywhere, Γ being rectifiable). Actually, these questions were left undecided in [7], pp. 144-148. (It was shown there, among other things, that every Γ must possess the "Archimedian property"; and while it is easy to see that this property of a rectifiable Jordan arc does not imply the existence of a tangent at every point of the arc,[1] it seems to be less easy to "embed" such an arc into a metric (2) for which the arc becomes a geodesic.) Fortunately, these matters will not complicate the situation, since the nature of all the problems considered is such as to imply the C^1-character of those particular (even though possibly not of all) geodesics Γ of a metric (2) to which the assertions of the theorems refer.

2. *Transversals.* The principal results will depend on (the wording or the proof of) a lemma concerning transversal fields; cf. [6], pp. 147-151. Under the classical assumptions of differentiability, the assertion of the lemma is nothing but a theorem of Jacobi, and is then a corollary of the theorem of Gauss concerning the transversal trajectories of a sheaf of geodesics (cf.

[1] In order to obtain such a Jordan arc in a (u, v)-plane, it is sufficient to put $u = r\cos\phi$, $v = r\sin\phi$, and to assign the Jordan arc in the parametric form $r = f(\phi)$, where $f(\phi)$, with $f(\phi_1) \neq f(\phi_2)$ when $\phi_1 \neq \phi_2$, is positive, is of class C^1 for $-\infty < \phi < \infty$ and tends sufficiently fast to 0 as $\phi \to \pm\infty$ (so that the point $(u, v) = (0, 0)$ of the Jordan arc belongs to $\phi = -\infty$ *and* $\phi = \infty$). In fact, an easy calculation shows that the choice $f(\phi) = \exp(-\phi^2)$ will do (the slower logarithmic spiral, $f(\phi) = \exp(-|\phi|)$, will not do; the non-differentiability of the latter function at $\phi = 0$ is of course immaterial).

[3], vol. II, p. 430). But the point is that, under the general assumption (*) below, the classical proofs must fail to apply. The lemma in question is as follows:

LEMMA 1. *Suppose that*

(0) *E, F, G in* (2) *are continuous functions satisfying* (3) *on a sufficiently small domain* (1), *and that*

there exists, on that domain, a function, say $\phi(u,v)$, which is of class C^1 and has the following properties:

(i) *all solutions of the differential equation*

$$dv/du = \phi(u,v) \tag{4}$$

represent geodesic arcs, Γ, of (2) *and*

(ii) *if a geodesic arc, Γ, of* (2) *possesses a tangent at some point and has, at that point, the same direction as a solution path of* (4), *then the latter solution path is identical with the arc Γ.*

Under these assumptions, there exist on (1) *positive, continuous multipliers $\mu = \mu(u,v)$ for the Pfaffian*

$$\omega = (E + F\phi)du + (F + G\phi)dv; \tag{5}$$

in fact, the function

$$\mu = (E + 2F\phi + G\phi^2)^{-\frac{1}{2}} \tag{6}$$

(*which, in view of* (0), *is continuous and positive*) *is such a multiplier.*

In other words, there exists on $\mathcal{B}_a$: $u^2 + v^2 < a^2$ a function $r = r(u,v)$ (unique to an additive constant, say to its value $r(0,0)$ at the center of $\mathcal{B}_a$) such that r possesses on $\mathcal{B}_a$ the partial derivatives

$$r_u = \mu(E + F\phi), \qquad r_v = \mu(F + G\phi), \tag{7}$$

where μ denotes the function (5).

Since the vanishing of the Pfaffian (5), i. e., the differential equation

$$(E + F\phi)du + (F + G\phi)dv = 0 \tag{8}$$

for $v = v(u)$ or $u = u(v)$, characterizes the transversals of the sheaf of curves defined by the differential equation

$$dv - \phi du = 0, \tag{9}$$

the assertion of Lemma 1 implies that the differential equation of the transversals of the geodesic sheaf defined by (4) possesses *some* (continuous, non-

vanishing) integrating factor, $\mu = \mu(u, v)$. Not even this is obvious, since, if nothing but (0) in Lemma 1 is assumed, and if (5) is written in the form $\omega = M(u, v)\,du + N(u, v)\,dv$, where M and N are just continuous, then the Pfaffian ω need not possess a multiplier in any sense; cf. [12], Section 7. Correspondingly (and in view of Section 7 below), it turns out that the content of Lemma 1 is substantially equivalent to the following statement:

LEMMA 2. *Under the assumptions of Lemma* 1 *(and if a in $\mathcal{B}_a$: $u^2 + v^2 < a^2$ is small enough), there exists on $\mathcal{B}_a$ a pair of functions,*

$$(10) \qquad u^* = u^*(u, v), \qquad v^* = v^*(u, v),$$

such that the transformation (10) *is of class C^1 and of non-vanishing Jacobian, and maps the neighborhood $\mathcal{B}_a$ of $(u, v) = (0, 0)$ on a neighborhood of $(u^*, v^*) = (0, 0)$ in such a way that the metric* (2) *acquires a "geodesic" form*

$$(11) \qquad ds^2 = du^{*2} + g(u^*, v^*)\,dv^{*2},$$

where g is a positive, continuous function.

Lemma 1 has a partial converse, as follows:

LEMMA 3. *Suppose that* (2) *satisfies* (0) *in Lemma* 1, *and that a function $\phi(u, v)$, which is of class C^1 on $\mathcal{B}_a$, has the property that the Pfaffian* (5) *possesses the multiplier* (6) *on $\mathcal{B}_a$. Then every solution of* (4) *represents a geodesic of* (2).

It remains undecided whether the C^1-assumption imposed on $\phi(u, v)$ in Lemmas 1-3 can be reduced in some extent (for instance, to the extent of requiring only that $\phi(u, v)$ be continuous and such that the solutions of (4), belonging to an initial condition, are unique).

3. *Reduction to parallel segments.* By adapting a scheme from analytical mechanics (in a highly differentiable case), it will be convenient to arrange the proofs in such a way that the sheaf of geodesics which is defined by (4) is assumed to be a sheaf of line segments

$$(12) \qquad v = \text{const.},$$

which means that

$$(13) \qquad \phi(u, v) \equiv 0$$

in (4). It turns out that it can be assumed that, besides (13),

$$(14) \qquad F(0, v) \equiv 0$$

in (2).

In order to make applicable the simplification afforded by this scheme, it will first be proved that, under the assumptions of Lemma 1, 2 or 3, there always exists an admissible (u, v)-transformation of class C^1 which leaves the assumptions unchanged and leads to the normalizations (13)-(14).

Proof. After an affine transformation of (1) (and if a in the new $\mathcal{B}_a$: $u^2 + v^2 < a^2$ is small enough), it can be assumed that the coefficient matrix of (2) is the unit matrix at the center of $\mathcal{B}_a$,

$$(15) \qquad E(0,0) = 1, \qquad F(0,0) = 0, \qquad G(0,0) = 1.$$

Then the function $E + F\phi$ of (u, v), being $1 + 0 \cdot \phi((0,0) > 0$ at the center of $\mathcal{B}_a$, will satisfy

$$(16) \qquad E + F\phi > 0 \text{ on } \mathcal{B}_a,$$

if a is small enough. Hence the differential equation (8) can be written in the form $u' = f(u, v)$, where f is continuous on $\mathcal{B}_a$ and the prime denotes d/dv. Consequently (8) has, for small $|v|$, say for $|v| < b$, at least one solution $u = u(v)$ satisfying $u(0) = 0$. Let $\gamma(v)$ denote such a solution of (8). Thus

$$(17) \qquad u = \gamma(v), \text{ where } -b < v < b \text{ and } \gamma(0) = 0.$$

On the other hand, since $\phi(u, v)$ is of class C^1, the differential equation (9), which is (4), has a unique solution $v = v(u) = v(u; u_0, v_0)$ satisfying $v(u_0) = v_0$, whenever (u_0, v_0) is close enough to $(0,0)$, and the function $v(u; u_0, v_0)$ exists and is of class C^1 in its three arguments together, if (u_0, v_0) is restricted to a sufficiently small circle about $(0,0)$, and u to a sufficiently short interval $-c < u < c$ (the length of which is independent of u_0, v_0).

In terms of the function $v(u; u_0, v_0)$ and of the function $\gamma(v)$ occurring in (17), define, for small $|\alpha|$, $|\beta|$, two functions, U and V, by placing

$$(18) \qquad U(\alpha, \beta) = \alpha + \gamma(\beta), \qquad V(\alpha, \beta) = v(\alpha + \gamma(\beta);\ \gamma(\beta), \beta),$$

and consider the transformation

$$(19) \qquad u = U(\alpha, \beta), \qquad v = V(\alpha, \beta).$$

This mapping of a neighborhood of $(\alpha, \beta) = (0,0)$ on a neighborhood of $(u, v) = (0,0)$ is of class C^1 (since the functions $\gamma(v)$, $v(u; u_0, v_0)$ occurring in (18) are), and the Jacobian of the transformation (19) does not vanish at, hence near, $(0,0)$. In fact, the Jacobian $\partial(U, V)/\partial(\alpha, \beta)$ of (18) is $V_\beta - \gamma'\phi$, where

$$V_\beta = \partial V/\partial\beta = \gamma'\phi + \gamma'\partial v/\partial u_0 + \partial v/\partial v_0.$$

Since $\partial v/\partial w_0$ at $(u; u_0, v_0) = (0; 0, 0)$ becomes 0 or 1 according as $w_0 = u_0$ or $w_0 = v_0$, it follows that the Jacobian at $(\alpha, \beta) = (0, 0)$ is $0 + 0 + 1 \neq 0$.

The meaning of the transformation (18)-(19) (which corresponds to the "transformation to the rectilinear motion" in the Hamilton-Jacobi theory) is as follows: For a given (u_0, v_0), the solution path $v = v(u; u_0, v_0)$ of (9) meets the solution path (17) of (8) at a unique point $(\gamma(\beta_0), \beta_0)$. The inverse of the transformation (18)-(19) is $(u_0, v_0) \rightarrow (\alpha_0, \beta_0)$, where $\alpha_0 = u_0 - \gamma(\beta_0)$. For a fixed β, the arc (19) is a solution path of (9).

It is seen either from this interpretation or from a direct calculation that, if (e, f, g) is the (α, β)-representation of the covariant tensor (E, F, G), i. e., if (2) is identical with

$$ds^2 = e(\alpha, \beta)\, d\alpha^2 + 2f(\alpha, \beta)\, d\alpha d\beta + g(\alpha, \beta)\, d\beta^2$$

by virtue of the C^1-transformations (18)-(19), then the functions e, f, g are identical with

$$2F\phi + G\phi^2, \qquad E\gamma' + F\gamma' + FV_\beta + G\phi V_\beta, \qquad E\gamma'^2 + 2F\gamma' V_\beta + GV_\beta^2,$$

respectively. It is clear from (18) that (4) is transformed into $d\beta/d\alpha = 0$, which means that (13) will hold if (α, β) is called (u, v). In addition, since the arc $u = \gamma(\beta)$, $v = \beta$ becomes the transversal (17) when $\alpha = 0$, condition (14) will be satisfied if $\alpha, \beta; f$ are called $u, v; F$. Finally, the Pfaffian (5) and the function (6) are transformed into $\omega^* = e d\alpha + f d\beta$ and $\mu^* = e^{-\frac{1}{2}}$, respectively. Since the property of a Pfaffian to be an exact differential and the property of an arc to be a geodesic are invariant under a local C^1-transformation $(u, v) \rightarrow (\alpha, \beta)$ of non-vanishing Jacobian, this proves that the assumptions (13) and (14) will not involve a loss of generality in the proof of Lemmas 1-3.

4. *Proof of Lemma* 1. In view of (13), the differential equations (9) and (8) reduce to

(20) $$dv/du = 0$$

and

(21) $$du/dv = -F(u, v)/E(u, v), \qquad (E > 0),$$

respectively, and since the function (6) becomes $\mu = E^{-\frac{1}{2}}$, the pair of relations (7) simplifies to

(22_1) $\quad r_u = E^{\frac{1}{2}}(u, v)$; $\qquad (22_2)$ $\quad r_v = F(u, v)/E^{\frac{1}{2}}(u, v)$.

Hence the assertion of Lemma 1 is that, under the assumptions (0) and (i)-(ii), there exists on $\mathcal{B}_a$ a function $r = r(u, v)$ of class C^1 satisfying both (22_1) and (22_2).

It will be proved that such an $r(u, v)$ is given by

$$(23) \qquad r(u, v) = \int_0^u E^{\frac{1}{2}}(t, v)\,dt.$$

It is clear that the function (23) is continuous and that its partial derivative $r_u(u, v)$ exists and satisfies (22_1) (and is therefore continuous). What is not clear is that the partial derivative $r_v(u, v)$ occurring in assertion (22_2) exists at all; in fact, the function integrated in (23) is just continuous.

With reference to a fixed number r, consider the equation

$$(24) \qquad r(u, v) = r\,(= \text{const.}).$$

In view of (23), this equation is satisfied at $(u, v) = (0, 0)$ if $r = 0$. It follows therefore from (22_1), where $E \neq 0$, and from standard facts on implicit functions, that (24) defines, in a neighborhood of $(v; r) = (0; 0)$, a unique continuous function

$$(25) \qquad u = u(v; r)$$

in such a way that a triple $(u, v; r)$ sufficiently close to the triple $(0, 0; 0)$ will satisfy (24) if and only if the u in $(u, v; r)$ is of the form (25). Furthermore, since $E \neq 0$ in (22_1), the function (25), defined by (24), has a continuous partial derivative with respect to r, and this derivative is given by

$$(26) \qquad u_r(u; r) = E^{-\frac{1}{2}}(u, v), \text{ where } u = u(v; r).$$

It will be proved that

(a) *there exists a positive number d having the property that, if r is any fixed value for which $|\,r\,|$ is sufficiently small, then the continuous function* (25) *exists for $|\,v\,| < d$ and represents a solution of* (21).

If (a) is granted, then the proof of Lemma 1 can be completed as follows: (a) implies that the function (25) has, in a neighborhood of $(v; r) = (0, 0)$, a partial derivative with respect to v, and that this derivative satisfies (21), i. e., that

$$(27) \qquad u_v(v; r) = -F(u, v)/E(u, v), \text{ where } u = u(v; r),$$

which implies that this derivative is continuous in v and r together. Since (24) and (25) are equivalent, it follows that the function (23) of (u, v) is of class C^1. Finally, (22_2) follows from (25), (26) and (27).

In order to prove (a), it will first be shown that

(β) *if a point* (u^0, v^0) *is sufficiently near the point* $(u, v) = (0, 0)$, *then the distance,* $\int ds$, *from the point* (u^0, v^0) *to the line* $u = 0$ *is minimized by the geodesic* $v = v^0$.

5. *Proof of* (β). This will be the only part of the proof of Lemma 1 in which the assumption (ii) is used.

First, if (u^0, v^0), where $u^0 \neq 0$, is a point close enough to the center of $\mathscr{B}_a$: $u^2 + v^2 < a^2$, then the existence of a rectifiable Jordan arc, say $\Gamma = \Gamma(u^0, v^0)$, minimizing the distance $(= \int ds)$ of (u^0, v^0) from the line $u = 0$, follows by the same procedure as the existence of an arc minimizing the distance between two given points which are close enough (Hilbert). If there are more than one $\Gamma = \Gamma(u^0, v^0)$, choose one of them, and denote by $(0, v^*)$ the point at which this Γ reaches the line $u = 0$. It will be shown that Γ is transversal to the line $u = 0$ at the point $(0, v^*)$. Then the assumption (ii) will assure that Γ is the geodesic $v = v^*$; cf. (12), (13). This in turn will imply that $v^* = v^0$ and that Γ is, therefore, the geodesic $v = v^0$, as claimed by (β).

Suppose if possible that $\Gamma = \Gamma(u^0, v^0)$ is not transversal to the line $u = 0$ at the point $(0, v^*)$. Then Γ either does not have a tangent at $(0, v^*)$ or, if it does, that tangent has a direction which fails to be transversal to $u = 0$ (by a tangent is meant a unilateral tangent, since Γ can be assumed to end at the point $v = v^*$ of the line $u = 0$). Under either of these hypothesis, it is seen from (14) and (21) that there exists on Γ a sequence of points (u_1, v_1), (u_2, v_2), $\cdots$ which converge to the point $(0, v^*)$ and have the property that the inequality

$$(28) \qquad |(v_n - v^*)/u_n| > c,$$

holds for a positive c which is independent of n $(= 1, 2, \cdots)$. Since $u^0 \neq 0$, it can be assumed that $u^0 > 0$, and also that (28) is replaced by

$$(29) \qquad v_n - v^* > cu_n > 0, \qquad (n = 1, 2, \cdots)$$

(the other possibilities can be treated similarly).

Accordingly, the proof of (β) will be complete if it is shown that the existence of a $c > 0$ satisfying (29), where $(u_n, v_n) \to (0, v^*)$ as $n \to \infty$, leads to a contradiction. The latter will result from the fact that (29) leads to the following conclusion:

(?) If n is large enough, then the distance $\int ds$ between the two points $(0, v_n)$, (u_n, v_n), when measured along the geodesic $v = v_n$, is shorter than

the distance $\int ds$ between the two points $(0, v^*)$, (u_n, v_n), when measured along $\Gamma = \Gamma(u^0, v^0)$.

In order to deduce the latter assertion, (?), subject the (u, v)-plane to an affine transformation $(u, v) \rightarrow (x, y)$ in such a way that (2) becomes $ds^2 = dx^2 + dy^2$ at the point $(u, v) = (0, v^*)$. Then the distance $\int ds$ along any geodesic which joins any two points is $(1 + o(1))d$ as those two points tend to the point $(0, v^*)$, where d denotes the Euclidean distance in the (x, y)-plane between the two points; cf. [7], pp. 144-148. On the other hand, it is clear that the Euclidean distance between the two points $(0, v^*)$, (u_n, v_n) is not less than $(1 + o(1))/\sin \alpha$ times the Euclidean distance between the two points $(0, v_n)$, (u_n, v_n), if, with reference to the positive number c occurring in (29), the $\alpha = \alpha_c$ in $1/\sin \alpha$, where $0 < \alpha < \frac{1}{2}\pi$, denotes the angle between the (x, y)-images of the two lines $v - v^* = cu$, $u = 0$. Since $1/\sin \alpha > 1$, and since $(u_n, v_n) \rightarrow (0, v^*)$ as $n \rightarrow \infty$, the preceding two o-relations imply the truth of (?).

The proof of (β) is now complete, since (?) contains a contradiction. In fact, if n is large enough, and if Γ_n denotes the path consisting of the portion $0 \leqq u \leqq u_n$ of the geodesic $v = v_n$ and of that portion of $\Gamma = \Gamma(u^0, v^0)$ which joins (u^0, v^0) to (u_n, v_n), then (?) implies that the length $\int ds$ of the path Γ_n, a path joining a point of the line $u = 0$ to the point (u^0, v^0), is shorter than the length $\int ds$ of $\Gamma(u^0, v^0)$. But this contradicts the definition of $\Gamma(u^0, v^0)$.

6. *Proof of* (α). Suppose if possible that the assertion of (α) is false. Then there exists on the arc (24) or (25) a point (u_0, v_0), arbitrarily close to the point $(0, 0)$, in such a way that the arc either does not have a tangent at (u_0, v_0) or, if it does, that tangent has a direction which fails to be transversal to the geodesic $v = v_0$. It can be assumed that $u \neq 0$, say $u_0 > 0$. For, if $r = 0$, then (23) shows that (25) is the arc $u = u(v; 0) \equiv 0$ which, in view of (14), represents a solution of (21).

In order to simplify the notations, subject the (u, v)-plane to an affine transformation after which the metric (2) reduces to $ds^2 = du^2 + dv^2$ at the point (u_0, v_0). Then, in both of the cases negating the truth of (α), there exists on the arc (24) or (25) a sequence of points $(u^1, v^1), (u^2, v^2), \cdots$ which converge to the point (u_0, v_0) and have the property that

$$(30) \qquad | v^n - v_0 | / | u^n - u_0 | < c, \text{ where } u^n \neq u_0 > 0,$$

holds for a positive c which is independent of $n(=1, 2, \cdots)$. Corresponding to the reduction of (28) to (29), it can be assumed that (30) is replaced by

$$0 \leqq v^n - v_0 < c(u_0 - u^n) \qquad (n = 1, 2, \cdots). \tag{31}$$

With reference to the constant $c > 0$ occurring in (31), let $\Lambda = \Lambda(u_0, v_0; c)$ denote the line $v - v_0 = (-1/c)(u_0 - u)$. Then it is clear from $(u^n, v^n) \to (u_0, v_0)$ that, if n is large enough, the line Λ will have on the geodesic $v = v^n$ some point, say the point (u_n, v^n).

If the results of [7], pp. 144-148, are applied in the same way as in Section **5** above, it now follows that if n is large enough, the distance $\int ds$ between the two points (u_0, v_0), (u_n, v^n), measured along the line Λ, is shorter than the distance $\int ds$ between the two points (u^n, v^n), (u_n, v^n), measured along the geodesic $v = v_n$. Hence there results a path consisting of a portion of the geodesic $v = v_0$ and of a segment of Λ and having the property that, while it is a path which joins the point $(0, v_0)$ to the point (u_n, v^n), it is shorter than the portion of the geodesic $v = v^n$ which joins $(0, v^n)$ to (u_n, v^n). This contradicts (β) and therefore proves (α).

The proof of Lemma 1 is now complete.

7. *Proof of Lemma* 2. In view of Section **3**, it is sufficient to prove Lemma 2 under the normalizations (13)-(15). Then, according to the proof of Lemma 1, the function (23) is of class C^1 and satisfies (22_1)-(22_2). In terms of this function, define a transformation of a neighborhood of $(u, v) = (0, 0)$ into a neighborhood of $(r, v) = (0, 0)$ by placing

$$r = r(u, v), \qquad v = v. \tag{32}$$

The Jacobian of the C^1-mapping (32) is $1 \cdot r_u - 0$, which, in view of (22_1), is distinct from 0.

Let the covariant metric tensor of (2) be expressed in terms of the coordinates (32). Then it is readily seen from (22_1)-(22_2) that

$$ds^2 = dr^2 + (EG - F^2)E^{-1}dv^2 \tag{33}$$

is an identity by virtue of (2) and (32). But (33) is of the desired form (11), with $u^* = r$, $v^* = v$, and $g = G - F^2/E$, where the functions E, F, G of u and v are thought of as expressed, by means of the inverse of the substitution (32), as functions of r and v.

8. *Proof of Lemma* 3. In view of Section **3**, it is sufficient to prove Lemma 3 under the assumptions (12)-(15). Then (14) and the assumptions

of Lemma 3 mean that (23) is a function of class C^1 satisfying (22_1)-(22_2).

Section 7 shows that (32) transforms (2) into (33). But (33) is of the form

$$ds^2 = dr^2 + g(r, v)\,dv^2, \qquad \text{where } g(0,0) = 1, \tag{34}$$

by (15), and it is clear from (34) that, if $r^2 + v^2$ is small enough, every segment (12) minimizes the distance $\int ds$ between any two points of that segment. This proves Lemma 3.

9. *Normal coordinates.* Suppose that a metric (2), satisfying (3), is given in terms of normal coordinates (u, v) at $(0, 0)$ (Riemann). By this is meant that

$$u = r\cos\theta, \qquad v = r\sin\theta \tag{35}$$

represents a geodesic for every fixed θ and that, in addition, the $r \geqq 0$ occurring in (35) is, except for a factor which depends only on θ, identical with the arc length on each of the geodesics $\theta = \text{const.}$ which issue from $(u, v) = (0, 0)$. *If the functions $E(u, v)$, $F(u, v)$, $G(u, v)$ possess partial derivatives of a sufficiently high order,* then, according to Gauss [5], pp. 249-250 (and, in the multi-dimensional case, Riemann [11], p. 279; cf. Dedekind's comments [4], pp. 406-407), the differential equations of the geodesics (that is, the equations of motion

$$[L]_u = 0, \qquad [L]_v = 0, \tag{36}$$

where the brackets denote the Lagrangian derivatives of

$$L(u, v; u', v') = \tfrac{1}{2}E(u, v)u'^2 + F(u, v)u'v' + \tfrac{1}{2}G(u, v)v'^2 = \tfrac{1}{2}s'^2 \tag{37}$$

and the prime denotes d/dt) must possess the invariant relations

$$L_{u'}(u, v; u, v) = L_{u'}(0, 0; u, v), \qquad L_{v'}(u, v; u, v) = L_{v'}(0, 0; u, v) \tag{38}$$

whenever (u, v) is a normal coordinate system at $(0, 0)$, and, as pointed out by Herglotz ([8], p. 216), this necessary condition for normal coordinates is sufficient as well.

Clearly, neither the definition of coordinates which are normal coordinates nor the pair of relations (38) (which, in view of (37), simply mean that

$$E(u, v)u + F(u, v)v = u, \qquad F(u, v)u + G(u, v)v = v, \tag{39}$$

if, without loss of generality, the normalization (15) is used) involves anything like the italicized hypothesis, that concerning a sufficiently high degree of smoothness of the coefficient functions of (2). Correspondingly, as a relevant

illustration of the above theory of geodesic fields in a metric which is just continuous, it will now be shown that the criterion holds without *any* differentiability assumption on the coefficient functions of (2).

(*) *Suppose that* $E(u,v)$, $F(u,v)$, $G(u,v)$ *are continuous functions satisfying* (3) *in a neighborhood of* $(0,0)$, *and that they are normalized at* $(0,0)$ *by* (15). *Then the pair of identities* (39) *is necessary and sufficient in order that the coordinates* u, v *be normal at* $(u,v)=(0,0)$.

The proof of the necessity and sufficiency of (39) could be deduced from Lemma 3 and Lemma 1, respectively. Corresponding to the circumstance that Lemmas 1-3 and (*) deal with "geodesic parallel" and "geodesic polar" coordinates, respectively, such a deduction of (*) would, however, lead to a complication, since what represents (4) in (*) is the differential equation $dv/du = v/u$, which has a singularity at $u=0$ (corresponding to the vanishing of the Jacobian of (35) at $r=0$). Because of this formal complication, it will be just as convenient to prove (*) directly.

10. *Proof of* (*). A direct substitution shows that (35) transforms (2) into

$$ds^2 = e(r,\theta)\,dr^2 + 2f(r,\theta)\,dr\,d\theta + g(r,\theta)\,d\theta^2, \tag{40}$$

where, in (binary) vector and matrix notations,

$$\begin{pmatrix} e \\ f/r \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} E & F \\ F & G \end{pmatrix} \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, \tag{41}$$

while

$$g/r^2 = E\sin^2\theta - 2F\sin\theta\cos\theta + G\cos^2\theta. \tag{42}$$

In order to prove the sufficiency of (39), suppose that the coordinates u, v satisfy (39). Then (35) shows that (41) can be written in the form

$$\begin{pmatrix} e \\ f/r \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

(where $1 = \cos^2\theta + \sin^2\theta$). Since this means that $e \equiv 1$ and $f \equiv 0$, it follows that (40) reduces to

$$ds^2 = dr^2 + g(r,\theta)\,d\theta^2 \qquad (g > 0). \tag{43}$$

But (43) makes it trivial that the distance $\int ds$ between the point $r=0$ and any point (r_0, θ_0), where $r_0 > 0$, is minimized by the path $\theta = \text{const.}\ (=\theta_0)$, and that r is the arc length along the path. This means that the coordinates (41) are normal at $(0,0)$, as claimed by the second assertion of (*).

In order to prove the necessity of (39), which is the first assertion of (*), it is sufficient to show that $e = 1$ and $f = 0$ are identities in (u, v) if the coordinates u, v are normal at $(0, 0)$. In fact, if the constants $e = 1$, $f = 0$ are substituted into (41), then (41) reduces to a pair of identities which, in view of the definition (35), is precisely (39).

According to the first line of (41) and the normalization (15), the value of $e(u, v)$ at $(u, v) = (0, 0)$ is $\cos^2 \theta + \sin^2 \theta = 1$. On the other hand, since the coordinates u, v are supposed to be normal at $(0, 0)$, the equations (35) represent, for every fixed θ, a geodesic on which the arc length, when measured from $(0, 0)$, is of the form $s = cr$, where the positive number $c = c(\theta)$ is independent of r. Hence it is clear from (40) that $e(u, v)$ is independent of (u, v). In view of $e(0, 0) = 1$, this proves that $e(u, v)$ is the constant 1.

Consequently, only the identical vanishing of f remains to be proved. But (40) shows that the identical vanishing of f is equivalent to the statement [2] that every arc $r =$ Const. is transversal to every geodesic $\theta =$ const. Suppose if possible that this transversality fails to take place somewhere, i. e., that there exist an $r_0 > 0$ and a $\theta = \theta_0$ for which the arc $r = r_0$ is not transversal to the geodesic $\theta = \theta_0$ at the point (r_0, θ_0). Then, by using results of [7], pp. 144-148, in the same way as above (Section 5), it follows that there exists a sequence of points (r_n, θ_n) which tend, as $n \to \infty$, to (r_0, θ_0) and possess the following property: Whenever n is large enough, the point (r_n, θ_n) can be joined to the origin $(r = 0)$ by an arc along which the length $\int ds$ is less than r_n. This contradicts, however, the assumption, according to which the arc $\theta = \theta_n$ joining the point (r_n, θ_n) to the point $r = 0$ is a geodesic arc of length r_n.

11. *Jacobi's multiplier.* While Lemmas 1-3 concern themselves with the equation (8) of the transversals of the geodesics defined by (9), the following lemma deals with (9) itself (but assumes that the function ϕ of u and v contains a parameter also).

LEMMA 4. *Let $\phi = \phi(u, v; w)$ be a function of class C^1 on the product space of a sufficiently small domain $\mathfrak{G}_a$: $u^2 + v^2 < a^2$ and of an interval $|w| < b$, and suppose that the functions E, F, G and ϕ satisfy assumptions* (0) *and* (i)-(ii) *of Lemma* 1, *when w is fixed. Then, for fixed w, the continuous function*

$$(44) \qquad \lambda = (EG - F^2)(E + 2F\phi + G\phi^2)^{-3/2}\phi_w$$

[2] This is the analogue of assertion (α), Section 4, in the proof of Lemma 1. Nothing like assertion (β), Section 4, is involved in the present case, since the arc $u = 0$, occurring in (β), now degenerates to the point $r = 0$.

of (u, v) represents a multiplier of (9). *What is somewhat more, there exists a function $R = R(u, v; w)$ which is continuous on the product space of $\mathcal{B}_a$ and $|w| < b$, is of class C^1 on $\mathcal{B}_a$, and satisfies the relations*

$$R_u = -\lambda\phi, \qquad R_v = \lambda. \tag{45}$$

(The vanishing of the partial derivative ϕ_w, and therefore that of the multiplier (44), *is not excluded.)*

If E, F, G, instead of being just continuous, are sufficiently smooth, then Lemma 4 reduces to Jacobi's theorem concerning a "last multiplier" (in case of geodesics); cf. [10], p. 498 and [3], vol. II, p. 431.

Lemma 4 will be essential in proving, without the usual assumptions of differentiability, a fundamental theorem on non-euclidean geometries; cf. Section 12 below.

Proof of Lemma 4. If w is fixed, then, according to assertions (5)-(6) of Lemma 1, the function

$$r(u, v; w) = \int_{(0,0)}^{(u,v)} \mu\{(E + F\phi)du + (F + G\phi)dv\} \tag{46}$$

satisfies (7) on $\mathcal{B}_a$. Since $\phi(u, v; w)$ is supposed to be of class C^1 (with the inclusion of w), the function (6) has the continuous partial derivative $\mu_w = -(F + G\phi)\mu^3\phi_w$. Hence, if (46) is differentiated with respect to w, a straightforward calculation gives

$$r_w(u, v; w) = \int_{(0,0)}^{(u,v)} (\lambda du - \lambda\phi dv), \tag{47}$$

if λ is defined by (44). Since (47) means that (45) is satisfied by the function

$$R(u, v; w) = r_w(u, v; w), \tag{48}$$

the assertions of Lemma 4 follow.

12. *Beltrami's theorem.* If all geodesic arcs of a metric (2) are segments of straight lines in a domain of some (u, v)-plane, then the metric is of constant curvature. This is a classical theorem of Beltrami ([1], pp. 262-280; cf. [3], vol. III, pp. 41-47). His proof and its variants assume, however, that the coefficient functions of (2) have a sufficiently high degree of differentiability in (u, v). Without such an assumption, the curvature K of

(2) cannot even be defined, and the proofs need substantially stronger restrictions of differentiability than what suffices for equating K to the differential operator assigned by the Theorema Egregium. On the other hand, in view of the fundamental *geometrical* significance of Beltrami's result, it seems to be essential that the theorem should be formulated and proved without any assumption of differentiability, as follows:

(**) *Suppose that the coefficients of* (2) *are continuous functions satisfying* (3) *on a* (u,v)*-domain, say on* $\mathfrak{B}_a$: $u^2+v^2<a^2$, *and that every sufficiently short segment* (*of a straight line,* $c_1u+c_2v=c_0$) *contained in* $\mathfrak{B}_a$ *is a geodesic arc of* (2). *Then the* (*given*) *coefficients of* (2) *are analytic functions of* (u,v) *on* $\mathfrak{B}_a$, *and the curvature* $K=K(u,v)$ *of* (2) *is independent of* (u,v) *on* $\mathfrak{B}_a$.

As mentioned in Section **1**, a point and a direction do not in general determine a (unique) geodesic of a metric (2), not even if the functions E, F, G of (u,v) are of class C^1 (which is not assumed in the present case). Correspondingly, a substantial part of the proof of the general formulation (**) of Beltrami's theorem will consist in showing that this and similar possibilities are excluded by the last assumption of (**). To this end, the following lemma will be needed:

(†) *Under the assumption of* (**), *every sufficiently short geodesic arc of* (2) *is a segment* (*of a straight line,* $c_1u+c_2v=c_0$) *in* $\mathfrak{B}_a$.

The proof of the latter assertion, (†), will depend on the steps used in the proof of Lemma 1.

13. *Proof of* (†). Let $\Gamma\colon (u=u(t), v=v(t))$, where $0\leqq t\leqq 1$, be a geodesic of (2) joining a point, say (u_0,v_0), of $\mathfrak{B}_a$ to another point of $\mathfrak{B}_a$. Without loss of generality, the latter point can be assumed to be the origin, $(0,0)$, since a is meant to be sufficiently small. The assertion of (†) is that the arc Γ must be on a straight line, (35), where $\theta=\theta(u_0,v_0)$ is constant on Γ, and $r=r(t)\geqq 0$, $0\leqq t\leqq 1$.

With reference to an interior point t^* of the given parameter range $0\leqq t\leqq 1$ of Γ, let Γ_1, Γ_2 denote the respective portions $0\leqq t\leqq t^*$, $t^*\leqq t\leqq 1$ of the given geodesic, $\Gamma=\Gamma_1+\Gamma_2$, and let Γ^1, Γ^2 be the segments (of straight lines) which join the point $P^*=(u(t^*),v(t^*))$ of Γ to its respective points $(u(0),v(0))=(0,0)$, $(u(1),v(1))$. Then Γ^i, where $i=1,2$, has the same length $\int ds$ as Γ_i. For, on the one hand, the assumptions of (†), being those of (**), imply that the segment Γ^i minimizes the distance $\int ds$ between its given end points and, on the other hand, the same is true of the geodesic

arc Γ_i, the end points of which are the same as those of Γ^i. Consequently, $\Gamma_1 + \Gamma^2$ and $\Gamma^1 + \Gamma_2$ are of the same length, and join the same points as the geodesic $\Gamma = \Gamma_1 + \Gamma_2$, as well as the polygonal path $\Gamma^1 + \Gamma^2$, and are therefor geodesics.

Since $\Gamma^1 + \Gamma^2$ is a geodesic containing the point $P^* = (u(t^*), v(t^*))$, an application of arguments applied in [7], pp. 144-148 (or, equivalently, in Section **5** above) shows that $\Gamma_1 + \Gamma^2$ must have at P^* a unilateral tangent line from the "right," and that this line is that containing the segment Γ^1. Since the latter determines for $\Gamma_1 + \Gamma^2$ a unilateral tangent line from the "left," it now follows from Corollary 1 in [7], p. 145, that $\Gamma_1 + \Gamma^2$ must have at P^* a tangent, i. e., that the two unilateral tangents coincide at P^*. For reasons of symmetry, the same is true of the geodesic $\Gamma^1 + \Gamma_2$, as well as of the geodesic $\Gamma^1 + \Gamma^2$.

Clearly, this is possible only if the two segments Γ^1, Γ^2 issuing from P^* are collinear, and if the straight line containing them represents a (bilateral) tangent of $\Gamma_1 + \Gamma_2 = \Gamma$ at P^*. Since the sum of the segments Γ^1, Γ^2 is a segment, $\Gamma^1 + \Gamma^2$ which is a geodesic joining the points $(u(0), v(0))$, $(u(1), v(1))$, points which are independent of the choice of $P^* = (u(t^*), v(t^*))$ on $\Gamma = \Gamma_1 + \Gamma_2$, it follows that Γ must be identical with the segment $\Gamma^1 + \Gamma^2$. This proves assertion (†) of Section **12**.

14. *Field constructions for* (**). Let w be any constant, and let

$$\phi = \phi(u, v; w) \equiv w \tag{49}$$

for every (u, v) on (1). The assumptions of (**) show that assumptions (0) and (i) of Lemma 1 are satisfied by the case (49) of (4) (for every fixed w). On the other hand, (†) shows that assumption (ii) of Lemma 1 is satisfied. Consequently, Lemma 1 is applicable (at every fixed value of w), as is Lemma 4 (the function (49) is of class C^1 in $u, v; w$ together).

Since the case $w = 0$ of (49) reduces (5) and (6) to $Edu + Fdv$ and $E^{-\frac{1}{2}}$ respectively, it follows from Lemma 1 that

$$E^{-\frac{1}{2}}(Edu + Fdv) \text{ is an exact differential.} \tag{50}$$

On the other hand, it is seen from the case (49) of the definition (44) of λ, and of the assertion (45) of Lemma 4, that since $\phi_w = w_w = 1$, the Pfaffian

$$(EG - F^2)(E + 2Fw + Gw^2)^{-3/2}(dv - wdu)$$

is an exact differential for every value of the constant w. Hence the choice $w = 0$ leads to the conclusion that

$$(EG - F^2)E^{-3/2} \text{ is a function of } v \text{ only.} \tag{51}$$

In addition, since the assumptions of (**), from which (50) and (51) have been concluded, remain unaltered if u, v and E, G are replaced by v, u and G, E respectively, it is clear that, corresponding to (50) and (51),

(52) $$G^{-\frac{1}{2}}(F\,du + G\,du) \text{ is an exact differential,}$$

and

(53) $$(EG - F^2)\,G^{-3/2} \text{ is a function of } u \text{ only.}$$

Beltrami's proof of the assertion, $K = \text{const.}$, of his theorem falls into two parts. He first concludes ([1], pp. 263-266) the preceding four relations, (50)-(53), under the assumption that the functions E, F, G are sufficiently smooth (of class C^n, with something like $n = 2$), and then, by assuming a still higher degree of differentiability (something like $n = 5$), he deduces ([1], pp. 266-270) the assertion, $K = \text{const.}$, from (50)-(53). Correspondingly, the completion of the proof (**) will depend on a suitable adaptation of the latter part of Beltrami's proof, leading from (50)-(53) to certain functional equations for which it turns out that their non-analytic solutions cannot be continuous (or, for that matter, L-integrable).

15. *The functional equations of* (**). Let $U = U(u)$, $V = V(v)$ denote, for sufficiently small $|u|$, $|v|$, the cube roots of the respective functions (53), (51). In view of (3), these continuous functions of the respective single variables u, v are positive. Since (53) and (51) mean that both products $UG^{\frac{1}{2}}$, $VE^{\frac{1}{2}}$ are identical with the cube root of $EG - F^2$, it follows that there exists a positive, continuous function $\lambda = \lambda(u, v)$ satisfying

(54) $$E^{\frac{1}{2}} = \lambda U, \qquad G^{\frac{1}{2}} = \lambda V,$$

and that a continuous function $\mu = \mu(u, v)$ is, therefore, defined by placing

(55) $$F = \mu\lambda UV$$

(the multipliers (44), (6) have nothing to do with the present λ, μ). In terms of (54) and (55), the remaining two relations, (50) and (52), mean that both Pfaffians

(56) $$\lambda U du + \mu V dv, \qquad \mu U du + \lambda V dv \quad \text{are exact differentials.}$$

Since U, V are positive, continuous functions of the respective single variables u, v, the conditions

(57) $$d\alpha = U du + V dv, \qquad d\beta = U du + V dv$$

and $\alpha(0,0) = 0, \beta(0,0) = 0$ define, near $(u,v) = (0,0)$, a pair of functions

$\alpha = \alpha(u, v)$, $\beta = \beta(u, v)$ which are of class C^1 and of non-vanishing Jacobian (the latter being $\partial(\alpha, \beta)/\partial(u, v) = -2UV < 0$). Thus $(u, v) \to (\alpha, \beta)$ and $(\alpha, \beta) \to (u, v)$ are one-to-one transformations, of class C^1, of corresponding small neighborhoods of the origins of the parameter planes (u, v), (α, β). Since $U(u)$ and $V(v)$ are positive and continuous, it follows that, in a neighborhood of $(\alpha, \beta) = (0, 0)$, two positive continuous functions, A and B, of single variables are defined by placing

$$(58) \qquad 1/U(u) = A(\alpha + \beta), \qquad 1/V(v) = B(\alpha - \beta).$$

In fact, since (57) means that $d(\alpha + \beta) = Udu$ and $d(\alpha - \beta) = Vdv$, the functions u, v of (α, β) depend only on $\alpha + \beta$, $\alpha - \beta$ respectively.

According to (56), both Pfaffians

$$(\lambda + \mu)(Udu + Vdv), \qquad (\lambda - \mu)(Udu - Vdv)$$

are exact differentials. In view of (57), this means that the same is true of both Pfaffians $(\lambda + \mu)\,d\alpha$, $(\lambda - \mu)\,d\beta$ (in which the functions λ, μ of (u, v) are thought of as expressed in terms of the new variables, α and β). Consequently, the continuous functions $\lambda + \mu$, $\lambda - \mu$ of (α, β) depend only on α, β respectively. Thus if $2a$, $2b$ denote these continuous functions of the respective single variables α, β, then

$$(59) \qquad \lambda = a(\alpha) + b(\beta), \qquad \mu = a(\alpha) - b(\beta).$$

Hence it is seen from the definitions of U and V, (54), (55) and (58), that

$$(60) \qquad a(\alpha) + b(\beta) = 4a(\alpha)b(\beta)A(\alpha + \beta)B(\alpha - \beta).$$

As pointed out above, λ and A, B are positive. It follows therefore from the first of the relations (59) and from (60) that $a > 0$, $b > 0$. In particular, division of (60) by $ab \neq 0$ is allowed; so that

$$(61) \qquad 1/a(\alpha) + 1/b(\beta) = 4A(\alpha + \beta)B(\alpha - \beta).$$

If $\alpha + \beta$ and $\alpha - \beta$ in (61) are replaced by 2α and 2β, respectively, then an integration with respect to the new β leads to the following identity in (α, β):

$$\int_{\alpha}^{\alpha+\beta} dt/a(t) - \int_{\alpha}^{\alpha-\beta} dt/b(t) = 4A(2\alpha)\int_{0}^{\beta} B(2t)\,dt.$$

Since the sum on the left of this identity has a continuous partial derivative with respect to α, the same is true of the product on the right. This means that $A(\alpha)$ has a continuous first derivative. If the rôles of the original α and

β are interchanged in this deduction, it follows that $B(\beta)$ has a continuous first derivative. Finally, a repetition of this argument shows that $A(\alpha)$ and $B(\beta)$ have derivatives of arbitrarily high order.

Consequently, the tacit assumptions of the caluculations of Beltrami ([1], pp. 266-272), referred to at the end of Section **14** above, are satisfied by necessity. In other words, the classical proofs of Beltrami's theorem can now be repeated; so that the proof of (**), Section **12**, is complete.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] E. Beltrami, *Opere Matematiche*, vol. 1 (1902).

[2] O. Bolza, *Vorlesungen über Variationsrechnung*, Leipzig und Berlin, 1909.

[3] G. Darboux, *Leçons sur la théorie générale des surfaces*, Paris, vol. II (1889) and vol. III (1894).

[4] R. Dedekind, see [11].

[5] C. F. Gauss, *Werke*, vol. 4 (1873).

[6] J. Hadamard, *Leçons sur le calcul des variations*, Paris, 1910.

[7] P. Hartman and A. Wintner, "On the problems of geodesics in the small," *American Journal of Mathematics*, vol. 73 (1951), pp. 132-148.

[8] G. Herglotz, Zur Riemannschen Metrik, Berichte der Sächsischen Akademie der Wissenschaften zu Leipzig, vol. 73 (1921), pp. 215-225.

[9] D. Hilbert, "Ueber das Dirichlet'sche Prinzip," *Jahresbericht der Deutschen Mathematiker-Vereinigung*, vol. 8 (1900), pp. 184-188.

[10] C. G. Jacobi, *Gesammelte Werke*, vol. 4 (1886).

[11] B. Riemann, *Gesammelte mathematische Werke* (ed. 1892).

[12] A. Wintner, "On isometric surfaces," *American Journal of Mathematics*, vol. 74 (1952), pp. 194-214.

NOTE ON DOUBLE-MODULES OVER ARBITRARY RINGS.*

By TADASI NAKAYAMA.

Jacobson's [3] module-theoretical Galois theory of non-normal extension fields was generalized by Hochschild [2] to a theory of double-modules over sfields. The theory was further extended to the infinite-dimensional case by Dieudonné [1]. On the other hand, it was shown in [4] that a goodly portion of the theory can be transferred to the case of general rings. It is now natural to study the infinite-dimensional case also for general rings, which we propose to do in this note. It seems to the writer that there are [1] two main features in the theory; one is the characterization of relation-modules, and the other is the characterization of direct (Kronecker) self-products. With respect to the latter, our generalization is rather satisfactory (§ 5), while it is quite powerless with respect to the former; the same was the case with [4]. In closing the introduction, we remark that the formulation of the present note is left-right symmetric to the one in [4] (but is in accord with [1], [2]).

1. Relation-modules of double-modules. Let K be a ring; by a ring we mean in this note always one which posesses a unit element, by a subring we mean one containing that unit element, and by a module, either left- or right-, we mean one on which the unit element operates as an identity. Let A be a second ring, and $\mathfrak{A}$ the additive group of all additive homomorphisms of A into K. $\mathfrak{A}$ is an A-K-double-module with respect to the natural operations defined by [2]

$$x(a \cdot \alpha) = (xa)\alpha, \qquad x(\alpha \cdot k) = (x\alpha)k \tag{1}$$
$$(x, a \in A;\, k \in K;\, \alpha \in \mathfrak{A}).$$

Let M be a K-A-double-module, and let M^* be the module of all K-homomorphisms of M into K, i. e. the dual module of the K-(left-)module M. It is an A-K-double-module according to a definition similar to the above.

* Received May 31, 1951.

[1] Besides the duality between certain subrings (over which the whole ring has independent (right-) bases) and certain relation-modules.

[2] The operations of the elements of A, K on $\mathfrak{A}$ are indicated by dots.

(In fact, the above construction depends only on the fact that A is an A-right-module and K is a K-right-module.) Let u_0 be an element of M. With $\sigma \varepsilon M^*$ we consider the element $\bar{\sigma}$ of $\mathfrak{A}$ defined by

$$x\bar{\sigma} = (u_0 x)\sigma \qquad (x \varepsilon A). \tag{2}$$

Denote the totality of $\bar{\sigma}$'s (σ running over M^*), by $R(M, u_0)$. It is an A-K-submodule of $\mathfrak{A}$ as is readily seen, and we call it the *relation-module* of u_0 in M. We have evidently

LEMMA 1. *If M is contained in a A-K-double-module M_1, as K-A-module, such that every element of M^* can be extended to an element of M_1^*, then $R(M_1, u_0) = R(M, u_0)$.*

Let N be a second K-A-module, and v_0 an element of N. Let ϕ be a (K-A-)homomorphism of M into N. It induces in a natural manner an (A-K-)homomorphism $\phi^* : \tau \to \sigma = \phi\tau$ of N^* into M^*. Suppose now $u_0\phi = v_0$. Let $\bar{\tau}$ in $R(N, v_0)$ be given by $x\bar{\tau} = (v_0 x)\tau$. Then

$$x\bar{\tau} = (v_0 x)\tau = ((u_0\phi)x)\tau = ((u_0 x)\phi)\tau = (u_0 x)\sigma = x\bar{\sigma}. \tag{3}$$

Thus we have the following proposition, whose latter half includes Lemma 1.

PROPOSITION 1. *If there exists a (K-A-)homomorphic mapping ϕ of M into N which maps u_0 into v_0, then*

$$R(M, u_0) \supseteq R(N, v_0). \tag{4}$$

If ϕ^* maps N^* onto M^* then $R(M, u_0) = R(N, v_0)$.

We can immediately verify

PROPOSITION 2. *If $u_0 \varepsilon M$, $v_0 \varepsilon N$, then*

$$R(M \oplus N, w_0) = R(M, u_0) + R(N, v_0) \tag{5}$$

with $w_0 = u_0 + v_0$ in $M \oplus N$.

Let S be a third ring, and let N be an S-K-double-module (contrary to the above). The direct product $N \times M = N \times_K M$ of N, M over K is defined as usual, and is an S-A-double-module. With $\sigma \varepsilon M^*$, $\tau \varepsilon N^*$ (where N^* is the dual module of the S-module N), we set

$$(v \times u)(\sigma \times \tau) = (v(u\sigma))\tau, \tag{6}$$

and observe that this essentially defines $\sigma \times \tau$ as an element of $(N \times M)^*$, independent of the special expression of $v \times u$. Let $\bar{\sigma}$, $\bar{\tau}$ be the elements of

$R(M, u_0)$, $R(N, v_0)$ (v_0 being an element of N), which correspond to σ, τ, respectively. We have

$$(x\bar{\sigma})\bar{\tau} = ((u_0x)\sigma)\bar{\tau} = (v_0((u_0x)\sigma))\tau$$
$$= (v_0 \times (u_0x))(\sigma \times \tau) = ((v_0 \times u_0)x)(\sigma \times \tau).$$

Hence $x \to (x\bar{\sigma})\bar{\tau}$ is the element of $R(N \times M, v_0 \times u_0)$ corresponding to $\sigma \times \tau \varepsilon (N \times M)^*$, and we have

PROPOSITION 3. *For the product $N \times M$,*

$$R(M, u_0)R(N, v_0) \subseteq R(N \times M, v_0 \times u_0). \tag{7}$$

Furthermore we have

PROPOSITION 4. *Let M be a K-A-double-module generated, as K-A-module, by a single element u_0. Let N be a second K-A-double-module such that for every non-zero element v of N, there exists at least one element τ of N^* satisfying $v\tau \neq 0$. Then the inclusion (4) implies, conversely, the existence of a (unique) homomorphism ϕ of M into N, such that $u_0\phi = v_0$.*

For a proof we observe, after Hochschild [2], that if a certain sum $\Sigma\, ku_0a$ ($k \varepsilon K, a \varepsilon A$) vanishes, then $\Sigma\, k(a\bar{\sigma}) = \Sigma\, k((u_0a)\sigma) = (\Sigma\, ku_0a)\sigma = 0$. If (4) is the case, then this implies $\Sigma\, k(a\bar{\tau}) = 0$, or $(\Sigma\, kv_0a)\tau = 0$, for all $\tau \varepsilon N^*$. This implies in turn $\Sigma\, kv_0a = 0$, according to our assumption on N^*. Thus $u_0 \to v_0$ defines a (K-A-)homomorphism of M into N.

2. Restricted relation-modules. So far, no particular assumptions have been made on M (nor on K, A). Let us now assume that M possesses an independent K-basis, say $\{u_h\}$. Let σ_h be the element of M^* such that

$$u_i\sigma_h = \delta_{ih} \qquad \text{(Kronecker } \delta\text{'s)}; \tag{8}$$

$u\sigma_h$ is nothing but the coefficient of u_h in the K-linear expression of u by $\{u_h\}$. The K-combinations of σ_h's (with varying h) form a K-submodule $M^\#$ of M^*, which is independent of the particular choice of the basis $\{u_h\}$. Letting σ run only over $M^\#$, we then obtain a submodule $R^\#(M, u_0)$ of $R(M, u_0)$, which we shall call the *restricted relation-module* of u_0 in M. If $\bar{\sigma}_h$ is obtained from σ_h, then $\{\bar{\sigma}_h\}$ forms a (not necessarily independent) K-basis of $R^\#(M, u_0)$. Unlike $R(M, u_0)$, the restricted relation-module $R^\#(M, u_0)$ is not A-left-allowable in general.

Similarly to Lemma 1 we have

LEMMA 2. *If M is contained in a K-A-double-module M_1 as K-A-module,*

and M_1 has an independent K-basis which contains an independent K-basis of M, then $R^{\#}(M_1, u_0) = R^{\#}(M, u_0)$.

Next let M be homomorphically mapped into N by ϕ, and $u_0\phi = v_0$. Suppose also N has an independent K-basis, say $\{v_j\}$. Expressing each $u_h\phi$ by $\{v_j\}$, we see that for $\tau \varepsilon N^{\#}$, the element $\phi\tau$ of M^* is in $M^{\#}$. It follows readily that $R^{\#}(N, v_0) \subseteq R^{\#}(M, u_0)$. That Proposition 2 can be transferred to restricted relation-modules is trivial. Thus

PROPOSITION 5. *Our Propositions* 1, 2 *hold also for the restricted relation-modules $R^{\#}$ (instead of R), provided that both M, N possess independent K-bases.*

As for Proposition 3, we observe that if $\{u_h\}$, $\{v_j\}$ are, respectively, an independent K-basis of M and an independent S-basis of N, then $\{v_j \times u_h\}$ is an independent S-basis of $N \times M = N \times_K M$. Also

$$(9) \qquad (v_j \times u_h)(\sigma_i \times \tau_k) = (v_j(u_h\sigma_i))\tau_k = \delta_{hi}\delta_{jk}.$$

On the other hand, the element of $R^{\#}(N \times M, v_0 \times u_0)$ corresponding to $\sigma_i \times \tau_k$, is just the product $\bar{\sigma}_i\bar{\tau}_k$, as was seen formerly. Thus we have

PROPOSITION 6. *Let M, N be K-A-, and S-K-double-modules, respectively, possessing independent K-, S-bases. Then*

$$(10) \qquad R^{\#}(M, u_0)R^{\#}(N, v_0) \supseteq R^{\#}(N \times M, v_0 \times u_0).$$

We have furthermore, corresponding to Proposition 4,

PROPOSITION 7. *Let M, N be K-A-double-modules possessing independent K-bases. If M is generated by u_0 as a K-A-module, and if $R^{\#}(M, u_0) \supseteq R^{\#}(N, u_0)$, then $u_0 \rightarrow v_0$ gives a (K-A-)homomorphism of M into N.*

Let us next consider a K-A-double-module M which (not only is generated by u_0 and possesses an independent K-basis, but) possesses an independent K-basis consisting of elements contained in u_0A; then we say that M is a *special K-A-double-module*, and u_0 is a generator. We prove

PROPOSITION 8. *If M is a special K-A-double-module with generator u_0, and $\{u_h\}$ is anarbitrary independent K-basis of M (not necessarily contained in u_0A), then $\{\bar{\sigma}_h\}$ (with σ_h as in* (8)) *forms an independent K-(right-)basis of $R^{\#}(M, u_0)$.*

It suffices to consider the case where $\{u_h\}$ is contained in u_0A. Let

$$(11) \qquad u_h = u_0 t_h \qquad (t_h \varepsilon A).$$

Then

(12) $$t_h\bar{\sigma}_i = \delta_{hi}.$$

It follows immediately that the $\bar{\sigma}_i$'s are independent. That they form a K-basis of $R^{\#}(M, u_0)$ has been seen before.

3. Direct self-products of rings. We now consider the case $K = A$. Then $\mathfrak{A}$ is the absolute module-endomorphism ring of A. Let S be a subring of A. The direct self-product $M = A \times {}_S A$ of A over S is an A-double-module (i. e. A-A-double-module), and we have

PROPOSITION 9. *The relation-module $R(M, 1 \times 1)$ of $M = A \times {}_S A$ with respect to 1×1, is the S-left-endomorphism ring of A (i. e. the commuter in $\mathfrak{A}$ of the left-multiplication ring S_L of S on A), denoted by $E(A, S_L)$.*

For, with $\bar{\sigma} \varepsilon R(M, 1 \times 1)$, where $\sigma \varepsilon M^*$, we have $x\bar{\sigma} = ((1 \times 1)x)\sigma = (1 \times x)\sigma$. For $s \varepsilon S$ we have

$$(sx)\bar{\sigma} = (1 \times sx)\sigma = (s \times x)\sigma = (s(1 \times x))\sigma = s((1 \times x)\sigma) = s(x\bar{\sigma}),$$

which shows that $\bar{\sigma} \varepsilon E(A, S_L)$. Conversely, if $\alpha \varepsilon E(A, S_L)$, we put $(x \times y)\sigma = x(y\alpha)$, and observe that this defines σ uniquely as an element of M^*, since $(xs \times y)\sigma = (xs)(y\alpha) = x(sy\alpha) = (x \times sy)\sigma$. Clearly $\alpha = \bar{\sigma}$.

Now let M be an arbitrary A-double-module and u_0 an element of M. Then the set of all x satisfying $au_0 = u_0a$ forms a subring of A, and we take it as S:

(13) $$S = \{x \varepsilon A \mid xu_0 = u_0x\}.$$

Then we have $R(M, u_0) \subseteq E(A, S_L)$; the proof is similar to the first half of the above proofs of Proposition 9 (which depended only on $(1 \times 1)s = s(1 \times 1)$). In other words, if $s \varepsilon S$, then the left-multiplication s_L (on A) commutes with every element of $R(M, u_0)$. The converse is also true if M is such that for every non-zero element u in M, there exist a σ in M^* satisfying $u\sigma \neq 0$.

Now we wish to know when M is isomorphic to $A \times {}_S A$. To do so, let us assume that M is special, and u_0 is its generator. Let $\{u_h\}$ be an independent A-left-basis of M contained in u_0A, and put $u_h = u_0t_h$ (i. e. (11)). We have then (12). Let N be a second (special) A-double-module which is isomorphic to M. Let v_0, v_h be its elements corresponding to u_0, u_h. Suppose that $u_0 \rightarrow v_0 \times u_0$ gives an (A-two-sided) homomorphism of M into $N \times M$. By the homomorphism, u_0x is mapped on $(v_0 \times u_0)x = \Sigma_h v_0 \times (x\bar{\sigma}_h)u_h$

$=\Sigma_{h,i}((x\bar{\sigma}_h)\bar{\sigma}_i)(v_i\times u_h)$. On the other hand, $u_0x=\Sigma_h(x\bar{\sigma}_h)u_h=\Sigma_h(x\bar{\sigma}_h)u_0t_h$, and this is mapped on

$$\Sigma_h(x\bar{\sigma}_h)(v_0\times u_0)t_h=\Sigma_h(x\bar{\sigma}_h)(v_0\times u_h)=\Sigma_{h,i}(x\bar{\sigma}_h)(1\bar{\sigma}_i)(v_i\times u_h).$$

Hence

$$(x\bar{\sigma}_h)\bar{\sigma}_i=(x\bar{\sigma}_h)(1\bar{\sigma}_i). \tag{14}$$

So,

$$u_0(x\bar{\sigma}_h)=\Sigma_i((x\bar{\sigma}_h)\bar{\sigma}_i)u_i=\Sigma_i(x\bar{\sigma}_h)(1\bar{\sigma}_i)u_i=(x\bar{\sigma}_h)u_0.$$

Hence $x\bar{\sigma}_h\in S$ (for every $x\in A$ and every h). Also

$$\Sigma_h u_0(x\bar{\sigma}_h)t_h=\Sigma_h(x\bar{\sigma}_h)u_0t_h=\Sigma_h(x\bar{\sigma}_h)u_h=u_0x.$$

If $\mathfrak{r}$ denotes the right-ideal $\{x\in A\mid u_0x=0\}$ of A, then

$$x-\Sigma_h(x\bar{\sigma}_h)t_h\in\mathfrak{r}. \tag{15}$$

It is clear, because of (12), that the t_h are S-left-independent modulo $\mathfrak{r}$. Hence $\{t_h\}$ forms an independent left S-basis of A mod. $\mathfrak{r}$. Suppose $\mathfrak{r}$ is 0. Then $x=\Sigma_h(x\bar{\sigma}_h)t_h$, and $\{t_h\}$ forms an independent left S-basis of A. It is then easy to verify, observing (12), (15) particularly, that M is (A-two-sided) isomorphic to $A\times{}_SA$, by the correspondence $\{u_h\}\to\{t_h\}$ over A.

Our assumption was that $u_0\to v_0\times u_0$ gives an (A-)homomorphism of M into $N\times M$. This, however, can be secured either by assuming $R(N\times M, v_0\times u_0)\subseteq R(M,u_0)$, or by assuming $R^{\#}(N\times M, v_0\times u_0)\subseteq R^{\#}(M,u_0)$, by Propositions 4, 7. This proves the second half of the following Proposition, whose first half is readily seen to be true.

Proposition 10. *Let M be a special A-double-module with generator u_0. Let N be a second special A-double-module which is isomorphic to M, and v_0 its generator corresponding to u_0. M is isomorphic to $A\times{}_SA$ for some subring S of A, with u_0 corresponding to 1×1, if and only if*

$$R(N\times M, v_0\times u_0)\subseteq R(M,u_0)\quad(\text{or } R^{\#}(N\times M, v_0\times u_0)\subseteq R^{\#}(M,u_0))$$

and $u_0x=0$ $(x\in A)$ implies $x=0$.

4. Topologies in M^*, $\mathfrak{A}$. Coming back to the general case where K and A are perhaps different, we consider K, A, and M all in their discrete topologies; the particular concern is in K. We then introduce in $\mathfrak{A}$ the weak topology, in which a neighborhood of 0 is the set of elements vanishing at a finite subset of A. We consider M^* also in its weak topology. The mapping $\sigma\to\bar{\sigma}$ of M^* onto $R(M,u_0)$ is then continuous, since $x_i\bar{\sigma}=0$ $(i=1,2,\cdots,m)$

are implied by $(u_0x_i)\sigma = 0$ $(i = 1, 2, \cdots, m)$. Provided that u_0 generates M, as K-A-module (i. e. $M = Ku_0A$), the mapping is open too, as we see readily; hence it is a homeomorphism, since it is 1-1 under the assumption. Further, *$R(M, u_0)$ is closed in $\mathfrak{A}$, whenever $M = Ku_0A$* (or, more generally, when every element of $(Ku_0A)^*$ can be extended to an element of M^*). To prove this, let α be an element of the closure of $R(M, u_0)$ in $\mathfrak{A}$. We then put

$$(\sum_{i=1}^{m} k_i u_0 x_i)\sigma = \sum_{i=1}^{m} k_i(x_i\alpha),$$

where m is an arbitrary natural number, and k_i, x_i are arbitrary elements of K, A. This gives a unique definition of σ as a mapping of M into K. For, with different expressions of an element of M as sums of elements of the form ku_0x, the right-hand sides are equal, since the same is the case with arbitrary α in $R(M, u_0)$. σ is clearly K-linear, and $\sigma \varepsilon M^*$. Furthermore, α is equal to the $\bar{\sigma}$ given by this σ, which proves our assertion.

Assume that M possesses an independent K-basis. Then ($M^\#$ is dense in M^* whence) *$R^\#(M, u_0)$ is dense in $R(M, u_0)$*. Propositions 3, 6, combined, give

PROPOSITION 11. *Under the same assumption as in Proposition 6, $R(N \times M, v_0 \times u_0)$ is the closure of*

$$R(M, u_0)R(N, v_0) \quad (\textit{or, of } R^\#(M, u_0)R^\#(N, v_0))$$

in $\mathfrak{A}$.

5. The main theorem. Consider again the case $K = A$, and consider a special A-double-module M with generator u_0. Suppose $R(M, u_0)$ is a ring. Then Proposition 11 implies that the condition $R(N \times M, v_0 \times u_0) \subseteq R(M, u_0)$ in Proposition 10 is fulfilled. Thus we have

THEOREM. *A special A-double-module M is isomorphic to $A \times {}_S A$, with some subring S of A, if and only if the relation-module $R(M, u_0)$ forms a ring and $u_0x = 0$ implies $x = 0$.*

6. Characterization of relation-modules. So far our theory has been fairly smooth. In particular, our Theorem generalizes the similar theorem in the sfield case. Coming back to the case $K \neq A$ and seeking for a characterization of relation-modules, we encounter difficulties (cf. the examples below), which do not prevail in the sfield case. However, before we give up, let us prove

LEMMA 3. *Suppose a K-right-submodule $\mathfrak{M}$ of (the A-K-module) $\mathfrak{A}$ has an independent K-basis $\{\alpha_h\}$ such that for each $x \varepsilon A$, almost all $x\alpha_h$ are 0. Suppose further that $\mathfrak{M}$ is dense in $A \cdot \mathfrak{M}$, and in fact that for each $a \varepsilon A$ and for a finite number of h, say $1, 2, \cdots, m$, almost all $a \cdot \alpha_i$ are contained in the closure of the K-right-module spanned by $\{\alpha_j \ (j \neq 1, 2, \cdots, m)\}$. Then $\mathfrak{M}$ is the restricted relation-module of a certain K-A-double-module.*

To prove the lemma, let $\mathfrak{M}^*$ be the dual-module of the K-(right-)module $\mathfrak{M}$, i. e. the K-left-module of all continuous K-homomorphisms of $\mathfrak{M}$ into K. Since $\mathfrak{M}$ is dense in $A \cdot \mathfrak{M}$, every element of $\mathfrak{M}^*$ can be considered as a continuous K-homomorphism of $A \cdot \mathfrak{M}$ into K, and $\mathfrak{M}^*$ is essentially the dual-module of $A \cdot \mathfrak{M}$. Hence $\mathfrak{M}^*$ is a K-A-double-module. Let u_h be the elements of $\mathfrak{M}^*$ such that $u_h\alpha_i = \delta_{hi}$, and let M be the K-(left-)submodule of $\mathfrak{M}^*$ spanned by these u_h. It is the totality of elements u of $\mathfrak{M}^*$ such that almost all $u\alpha_i$ are 0 (for each u). Let u_0 be the elements of $\mathfrak{M}^*$ defined by $u_0\alpha = 1\alpha$ $(\alpha \varepsilon \mathfrak{M})$, 1 being the unit element of A. Because of our assumption, $u_0\alpha_h = 0$ for almost all h. Hence $u_0 \varepsilon M$. Further, M is A-right-allowable. To prove this, let $u \varepsilon M$, $a \varepsilon A$. Define ua by $(ua)\alpha = u(a \cdot \alpha)$. Let $\{1, 2, \cdots, m\}$ be the totality of indices h such that $u\alpha_h \neq 0$. Almost all $a \cdot \alpha_i$ are in the closure of the K-module spanned by $\{\alpha_j \ (j \neq 1, 2, \cdots, m)\}$. Hence almost all $(ua)\alpha_i$ $(= u(a \cdot \alpha_i))$ are 0, which proves $ua \varepsilon M$. It is now easy to see that $\mathfrak{M}$ can be considered as the restricted relation-module of this K-A-module M; observe

$$(16) \qquad (u_0x)\alpha = u_0(x \cdot \alpha) = 1(x \cdot \alpha) = x\alpha \qquad (\alpha \varepsilon \mathfrak{M}).$$

PROPOSITION 12. *In order that a K-submodule $\mathfrak{M}$ of A is a restricted relation-module of a special K-A-double-module, it is (necessary and) sufficient that $\mathfrak{M}$ be dense in $A \cdot \mathfrak{M}$ and possess an (independent) K-basis $\{\alpha_h\}$ such that for each $a \varepsilon A$, almost all $a\alpha_h$ are 0, and there exist t_h in A satisfying $t_h\alpha_i = \delta_{hi}$.*

To prove this, we consider an arbitrary element a of A, and any finite number of h, say $1, 2, \cdots, m$. We want to show that almost all $a \cdot \alpha_i$ are in the closure of the K-module spanned by $\{\alpha_j \ (j \neq 1, 2, \cdots, m)\}$. Each $a \cdot \alpha_i$ is anyway in the closure of $\mathfrak{M}$, by assumption. As is easily seen, it is an infinite sum $\Sigma\, \alpha_h k_h^{(i)}$ $(k_h \varepsilon K)$ in the sense of convergent sum in the topology of $\mathfrak{A}$. Thus $t_h(a \cdot \alpha_i) = k_h^{(i)}$. On the other hand, $t_h(a \cdot \alpha_i) = (t_h a)\alpha_i$, and with a given h this is 0 except for a finite number of i, which proves our assertion. Thus $\mathfrak{M}$ is the restricted relation-module of the module M constructed in the proof of Lemma 3. Also $(u_0 t_i)\alpha_j = t_i\alpha_j = \delta_{ij}$. Hence $u_i = u_0 t_i$. It is now clear that M is a special K-A-double-module with generator u_0. (The necessity assertion in the proposition is evident.)

Further, we see readily

PROPOSITION 13. *In order that an A-K-submodule of $\mathfrak{A}$ be a relation-module of a special K-A-double-module, it is necessary and sufficient that it be closed in $\mathfrak{A}$ and possess a dense subset $\{\alpha_k\}$ such that for every $a \in A$, almost all $a\alpha_k$ are 0, and there exist elements t_k in A satisfying $t_k\alpha_i = \delta_{ki}$.*

Remark. The condition of the existence of t_k in our Propositions is very strong, and in that sense our criteria of relation-modules are very poor. If, on the other hand, K is a sfield, then the automatic existence of α_k and t_k can be proved at least for K-finite modules (and also for K-infinite modules under a suitable topological condition) ([1], [2]; cf. also § 6 below), and thus we obtain a nice theorem in that case. However, such is certainly not the case in general, and our assumption seems indispensable,[3] as the following examples show:

Example 1. Let F be an arbitrary field and K the simple algebra of all 3-dimensional matrices over F. We take A to be identical with K. Let $e_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$. Let α_1 be the identity mapping of K $(=A)$, and α_2 the mapping $x \to e_1x$ $(x \in K)$. It is easy to see that α_1, α_2 are K-right-independent. The module $\alpha_1K \oplus \alpha_2K$ $(\subseteq \mathfrak{A})$ even forms a ring; the ring property is naturally of interest in connection with our theorem (in § 5). In spite of these nice properties of having an independent finite K-basis and being a ring, our module α-$K \oplus \alpha_2K$ does not possess the property required in Proposition 12 (or Proposition 13),[4] as we see without difficulty; the argument is similar to that in the next example.

Example 2. We now give an example in which K is a (non-commutative) integrity-domain. Let K be the integral domain in a $\mathfrak{p}$-adic division-algebra which is of degree 2 over its center. Let π be a primitive element of $\mathfrak{p}$ such that π^2 belongs to the center. We again take $A = K$. Let α_1 be the identity mapping of K $(=A)$, and α_2 the mapping $x \to \pi x$ $(x \in K)$. Again α_1, α_2 are right-independent over K, and $\alpha_1K \oplus \alpha_2K$ is a ring. Nevertheless, this module does not fulfill the requirement of Proposition 12 (or Proposition 13). (For, if there were $\beta_1, \beta_2 \in \alpha_1K \oplus \alpha_2K$ and $t_1, t_2 \in K$ with $t_k\beta_i = \delta_{ki}$, then the matrix $(t_k\alpha_i)_{ki}$ would be regular. But $\begin{pmatrix} t_2\alpha_1 & t_2\alpha_2 \\ t_1\alpha_1 & t_1\alpha_2 \end{pmatrix} = \begin{pmatrix} t_1 & t_1\pi \\ t_2 & t_2\pi \end{pmatrix}$, and this can not possess an inverse (in the 2-dimensional matrix ring over K)).

[3] It is very desirable, however, to find a simpler substitute.

[4] Or, what amounts to the same, K is not a regular module of $\alpha_1K \oplus \alpha_2K$ in the sense of [4].

7. Remarks on the case where K is a sfield. If K is in particular a sfield, then naturally every K-left-module M has an independent K-basis. Moreover, if N is a K-submodule of M, then every element of N^* can be extended to an element of M^*. Hence, as Lemmas 1, 2 (in § 1, 2) show, we may restrict ourselves, without loss in generality, to K-A-double-modules M which are generated by u_0 (i. e. cyclic modules in the sense of [1], [2]). Further, such K-A-modules are all special with generator u_0 (in our sense). Thus, except for the criteria (Propositions 12, 13) of relation-modules, our results seem to offer satisfactory generalizations of corresponding theorems in the sfield case. As for Propositions 12, 13, they do not include the perfect characterization of relation-modules given in [2], Theorem 5. 3 (finite-dimensional case) and [1], Theorem 1 (infinite-dimensional case); the pathological situation in the general case being exhibited in our examples in § 5 (they show that the immediate extension of [2], Theorem 5. 3 (or [1], Theorem 1) cannot be true). It is thus necessary to clarify the relationship between Proposition 13, for instance, and [1], Theorem 1. In fact, we prove

LEMMA 4. *If K is a sfield and if an A-K-submodule of A is linearly compact (over K) then it has the property of Proposition* 13.

To show this, we have actually to borrow the argument of [1], Theorem 1. Let M be the K-A-double-module consisting of all continuous K-homomorphisms of our module into K. Let u_0 be the element of M which sends every element of our module to its value at 1 (εA). Then $M = Ku_0A$, as was shown in [1]. Let $\{u_h\}$ be an independent K-basis of M contained in u_0A, and let $u_h = u_0t_h$. Let M^* be the dual module of the K-left-module M. It is essentially our given module.[5] (The dual of the dual of a linearly compact codule is, essentially, the module itself.) In fact, if we associate with each σ in M^* the element $\bar{\sigma}$ (as defined in § 1) of $\mathfrak{A}$, then the totality of $\bar{\sigma}$'s is exactly our original submodule of $\mathfrak{A}$. Now, let σ_h and $\bar{\sigma}_h$ be as in § 3. Then $\{\sigma_h\}$ is clearly dense in M^*, and therefore $\{\bar{\sigma}_h\}$ is dense in our module. But $t_i\bar{\sigma}_k = \delta_{ik}$. That almost all $a\bar{\sigma}_h$ are 0 for each a, is clear from the fact that $a\bar{\sigma}_h$ is simply the coefficient of u_h in the K-linear expression of u_0a by $\{u_h\}$.

(This lemma explains the relationship between Proposition 13 and [1], Theorem 1. However, since the essential part of the latter is used in the proof of our lemma, we do not claim, in the least, that the latter is contained in the former, as we want to repeat. (Moreover, in our proof of the lemma

[5] However, we prefer not to identify the two.

we actually proved that our module is a relation-module, and after having done so, Proposition 13 is rather superfluous.) In turn, Theorem 1 of [1] depends essentially on (the properties of linearly compact spaces, such as having discrete duals and being the duals of their duals, and) [2], Lemma 2. 1 (i. e. [1], Lemma 1), which is also essentially the theorem of dual bases of (finite) modules over a sfield).

UNIVERSITY OF ILLINOIS.

REFERENCES.

[1] J. Dieudonné, "Linearly compact spaces and double vector spaces over sfields," *American Journal of Mathematics*, vol. 73 (1951), pp. 13-19.

[2] G. Hochschild, "Double vector spaces over division rings," *American Journal of Mathematics*, vol. 71 (1949), pp. 443-460.

[3] N. Jacobson, "An extension of Galois theory to non-normal and non-separable fields," *American Journal of Mathematics*, vol. 69 (1947), pp. 27-36.

[4] T. Nakayama, "Non-normal Galois theory for non-commutative and non-semisimple rings," *Canadian Journal of Mathematics* (1951), pp. 208-218.

ORDER AND TOPOLOGY IN PROJECTIVE PLANES.*

By Oswald Wyler.

1. Introduction. In a finite-dimensional projective geometry with a topological coordinate field or skew-field, the topology of the coordinate field induces in a natural way a topology of the geometry, in which the line is closed and homeomorphic to the coordinate field with an additional element at infinity. This leads to the qeustion of whether a topology of a nondesarguesian plane can be deduced in a similar manner from a topology of its lines, and whether two such planes with homeomorphic lines will be homeomorphic.

In the most general case, this seems to be impossible without complicated additional assumptions. There are, however, two important exceptions to this statement, ordered projective planes, and planes with coordinates in an alternative division ring. In both cases the answer to our question is affirmative.

The first case is discussed in this paper. No additional assumption is needed beyond the axioms of incidence and of order.

Ordered projective planes have been studied by many authors,[1] chiefly in axiomatic treatments of real projective geometry. In this paper, a topology of an ordered projective plane is deduced from the interval topology on its lines, and the main properties of this topology of the plane are investigated. The Theorem of Desargues is never needed.

In this topology, the projective operations of joining two points by a line and of intersecting two lines in a point are continuous, and two planes with homeomorphic intervals are homeomorphic. This completes the answer to the question asked above. The homeomorphism between the intervals is assumed to map endpoints on endpoints. An order-preserving mapping of an interval onto another is such a homeomorphism. We do not have to assume, however, that our homeomorphism is order-preserving.

An example of a nondesarguesian ordered plane, given in section **9**, shows that the axioms of order, even with the Dedekind continuity axiom added, still make it necessary to assume a configuration theorem in order to get coordinates.

The decomposition of an ordered plane into three convex quadrangles

* Received June 22, 1951.

[1] Cf. e.g. [1], [6], [7]. Numbers in brackets refer to the references at the end of the paper.

with the same vertices, studied in section **6**, is constantly used in the next two sections.[2] It is much more useful for our purpose than the usual decomposition of a plane into four convex triangles with the same vertices.

Points are denoted by lower case letters, lines by small gothic letters. The line joining the points a and b is denoted by ab, the point of intersection of the lines $\mathfrak{l}$ and $\mathfrak{m}$ by $\mathfrak{l} \cap \mathfrak{m}$.

2. The axioms of separation. An *ordered projective plane* is characterized by the axioms of incidence and by a relation of *separation* between pairs of points, satisfying the axioms stated below. We write $ab \parallel cd$ if two points a and b separate two points c and d. The following axioms of separation are assumed.

S. 1. If $ab \parallel cd$, then a, b, c, d are four different collinear points.

S. 2. If a, b, c, d are four different collinear points, then at least one of the relations $ab \parallel cd$, $ac \parallel bd$, $bc \parallel ad$ holds.

S. 3. If $ab \parallel cd$, then $ab \parallel dc$.

S. 4. If $ab \parallel cd$ and $bc \parallel de$, then $cd \parallel ea$.

S. 5. If $ab \parallel cd$, and if a, b, c, d are mapped on a', b', c', d' by a perspectivity, then $a'b' \parallel c'd'$.

$ab \parallel cd$ always implies $cd \parallel ab$ by S. 1 and S. 5, since $abcd$ and $cdab$ are projective for any four different collinear points. This with S. 3 shows that separation is a symmetric relation between unordered pairs of points.

At most one of the relations in S. 2 can hold at the same time, for otherwise S. 4 would lead to a contradiction to S. 1.

If a line contains only three points, $ab \parallel cd$ never holds. If we assume that there is a line with four different points, then S. 2 and S. 5 imply

S. 6. If a, b, c are three different collinear points, then there is a point d such that $ab \parallel cd$.

Applying S. 6 repeatedly together with the other axioms, we see that an ordered projective plane cannot be finite.

Our axioms of separation are essentially those of [1], p. 22. Axiom S. 4 has been put into a more suggestive form.

---

[2] Cf. [2], p. 421, where this decomposition is denoted by $\{4, 3\}/2$ (for the elliptic plane). The author is indebted to the referee for this reference.

3. Segments and intervals. If a, b, c are three different points on a line $\mathfrak{g}$, we denote by $(ab)_c$ the set of all points x with $ab \parallel cx$. A set of this form is called a *segment* on $\mathfrak{g}$. The points a and b are called the *endpoints* of the segment $(ab)_c$. The set consisting of the segment $(ab)_c$ and its endpoints a and b is denoted by $[ab]_c$. Such a set is called an *interval* on $\mathfrak{g}$. We recall some well known properties of segments.

If $ab \parallel cd$, then the segments $(ab)_c$ and $(ab)_d$ are disjoint, and every point on $\mathfrak{g}$ different from the endpoints a and b is in one of them. If t is in $(ab)_d$, then $(ab)_t = (ab)_c$. There are thus exactly two segments with given endpoints a and b on $\mathfrak{g}$. A point in one of them shall be called an *exterior* point of the other.

If p and q are in $(ab)_c$, then $(ap)_c$ and $(pq)_c$ are contained in $(ab)_c$, and b is an exterior point of $(ap)_c$ and of $(pq)_c$. The first part of this proposition remains true if we replace segments by intervals.

LEMMA 3.1. *For any point u of a segment $(ab)_c$, there exist points p and q in $(ab)_c$, such that $(pq)_c$ contains u and is contained in $(ab)_c$. Then a and b are exterior points of $(pq)_c$.*

Proof. By S. 6 there is a point p such that $ua \parallel bp$. Then $cu \parallel ab$ implies $ab \parallel pc$ and $bp \parallel cu$ by S. 4. Thus p is in $(ab)_c$, and $(bp)_c$ contains u and is continued in $(ab)_c$. Similarly there is a point q in $(bp)_c$, such that $(pq)_c$ contains u and is contained in $(bp)_c$.

LEMMA 3.2. *The intersection of two segments $(ab)_t$ and $(cd)_t$ with a common exterior point t is either empty or a segment $(pq)_t$, where p and q are two of the endpoints of the given segments.*

Proof. If u is in both segments, then c is in one of the segments $(tu)_a$ and $(tu)_b$, and d in the other. If c is in $(tu)_b$ and d in $(tu)_a$, then $c = a$ or $ct \parallel ua$ or $at \parallel uc$ by S. 2, and similarly $d = b$ or $dt \parallel ub$ or $bt \parallel ud$. If the segments are not equal we may assume $ct \parallel ua$. Then $tu \parallel ab$ implies $ab \parallel ct$ by S. 4, hence c is in $(ab)_t$. If $b = d$, then $(cd)_t$ is contained in $(ab)_t$. This holds also for $dt \parallel ub$, since then d is in $(ab)_t$. If $bt \parallel ud$, then b is in $(cd)_t$; hence $(bc)_t$ is contained in $(ab)_t$ and in $(cd)_t$, and it is easily shown to be their intersection.

THEOREM 3.3. *If two segments $\mathfrak{a}$ and $\mathfrak{b}$ on a line $\mathfrak{g}$ have a point u in common, then there exists a segment on $\mathfrak{g}$ containing u and contained in the intersection of $\mathfrak{a}$ and $\mathfrak{b}$.*

Proof. Let s be an exterior point of $\mathfrak{a}$, t an exterior point of $\mathfrak{b}$. If $s = t$, then $\mathfrak{a} \cap \mathfrak{b}$ is a segment by Lemma 3.2. If $s \neq t$, then there exists a seg-

ment $\mathfrak{c}$ containing u with s and t as exterior points by Lemma 3.1. Then $\mathfrak{a} \cap \mathfrak{c}$ is a segment with exterior points s and t by Lemma 3.2, and hence $\mathfrak{a} \cap \mathfrak{c} \cap \mathfrak{b}$ is a segment containing u.

Theorem 3.3 shows that segments on a line $\mathfrak{g}$ define a topology on $\mathfrak{g}$ in which they are a base of open sets. In this topology the closure of the segment $(ab)_c$ is the interval $[ab]_c$, so that intervals form a base of closed neighborhoods. The topology is called the *interval topology* on $\mathfrak{g}$. By Lemma 3.1, $\mathfrak{g}$ is a regular space in its interval topology.

By S.5 a projectivity between two lines maps segments on segments and intervals on intervals, hence it is a homeomorphism between the two lines.

4. Sectors. By specifying that two pairs of lines in a pencil separate each other if and only if they intersect a line not in the pencil in pairs of points separating each other, we obtain a relation of separation for pairs of lines. Because of S.5, this definition is independent of the choice of the auxiliary transversal line. It is immediately seen that the dual propositions to our axioms S.1 to S.5 hold. Hence the definitions and results of the preceding section can be dualized. The dual of a segment will be called an *angle*.

Let $\mathfrak{l}$ and $\mathfrak{m}$ be two lines with a common point r, and let a and b be two points different from r. Then we shall say that $\mathfrak{l}\mathfrak{m} \parallel ab$ whenever $\mathfrak{l}\mathfrak{m} \parallel \mathfrak{p}\mathfrak{q}$ for $\mathfrak{p} = ra$ and $\mathfrak{q} = rb$. If $\mathfrak{l}$ and $\mathfrak{m}$ are two different lines, and if c is a point neither on $\mathfrak{l}$ nor on $\mathfrak{m}$, then we denote by $(\mathfrak{l};\mathfrak{m})_c$ the set of all points x of the plane such that $\mathfrak{l}\mathfrak{m} \parallel cx$. A set of this type is called a *sector*. The lines $\mathfrak{l}$ and $\mathfrak{m}$ are called the *sides*, and the point $\mathfrak{l} \cap \mathfrak{m}$ is called the *vertex* of the sector $(\mathfrak{l};\mathfrak{m})_c$.

To the sector $(\mathfrak{l};\mathfrak{m})_c$ corresponds the angle $(\mathfrak{l}\mathfrak{m})_{rc}$, where $r = \mathfrak{l} \cap \mathfrak{m}$, consisting of all lines of the pencil with vertex r through points of $(\mathfrak{l};\mathfrak{m})_c$. Conversely, the sector $(\mathfrak{l};\mathfrak{m})_c$ consists of all points on lines of the angle $(\mathfrak{l}\mathfrak{m})_{rc}$ and different from its vertex r, or in other words, of all points lying on exactly one line of the angle $(\mathfrak{l}\mathfrak{m})_{rc}$.

Thus the dual of a sector is the set of all lines meeting a given segment and different from the line containing the segment, or in other words, the set of all lines intersecting a given segment in exactly one point.

The following properties of sectors are immediately derived from the corresponding properties of segments on a line and of angles in a pencil.

Let $\mathfrak{l}$ and $\mathfrak{m}$ be two different lines with point of intersection r, and let c and d be two points such that $\mathfrak{l}\mathfrak{m} \parallel cd$. Then the sectors $(\mathfrak{l};\mathfrak{m})_c$ and $(\mathfrak{l};\mathfrak{m})_d$ are disjoint, and every point of the plane not on $\mathfrak{l}$ or on $\mathfrak{m}$ lies in one of them.

If t is in $(\mathfrak{l};\mathfrak{m})_d$, then $(\mathfrak{l};\mathfrak{m})_o = (\mathfrak{l};\mathfrak{m})_t$. There are thus exactly two sectors with given sides $\mathfrak{l}$ and $\mathfrak{m}$. A point in one of these sectors shall be called an *exterior* point of the other.

If $\mathfrak{g}$ is a line not through r, let $p = \mathfrak{g} \cap \mathfrak{l}$ and $q = \mathfrak{g} \cap \mathfrak{m}$. Then one of the segments with endpoints p and q is contained in $(\mathfrak{l};\mathfrak{m})_o$, the other in $(\mathfrak{l};\mathfrak{m})_d$. If $\mathfrak{g}$ is a line through the vertex r, different from $\mathfrak{l}$ and from $\mathfrak{m}$, then all points of $\mathfrak{g}$ different from r lie in the same sector with sides $\mathfrak{l}$ and $\mathfrak{m}$.

This implies immediately the following lemma.

LEMMA 4.1. *Let p and q be two different points of the sector $(\mathfrak{l};\mathfrak{m})_o$, and let $l = pq \cap (\mathfrak{l} \cap \mathfrak{m})c$. Then the segment $(pq)_t$ is contained in the sector $(\mathfrak{l};\mathfrak{m})_c$.*

5. The order topology of the plane. A set $\mathfrak{a}$ of points in the plane is called $\mathfrak{g}$-*convex*, where $\mathfrak{g}$ is a line, if $\mathfrak{a}$ and $\mathfrak{g}$ are disjoint, and if for any two different points p, q in $\mathfrak{a}$, the segment $(pq)_t$ is contained in $\mathfrak{a}$ for $t = pq \cap \mathfrak{g}$. A set of points is called *convex* if it is $\mathfrak{g}$-convex for some line $\mathfrak{g}$. A convex set is called *open* if its intersection with any line is an open set in the interval topology of the line.

A set consisting of one point not on $\mathfrak{g}$ is $\mathfrak{g}$-convex, and so is $\pi - \mathfrak{g}$, where π is the plane. Moreover, $\pi - \mathfrak{g}$ is open. If a and b are two different points not on $\mathfrak{g}$, and if $c = ab \cap \mathfrak{g}$, then the segment $(ab)_c$ and the interval $[ab]_c$ are $\mathfrak{g}$-convex. A sector $(\mathfrak{l};\mathfrak{m})_c$ is (rc)-convex and open for $r = \mathfrak{l} \cap \mathfrak{m}$. By Lemma 5.1 below, $(\mathfrak{l};\mathfrak{m})_c$ is also $\mathfrak{l}$-convex and $\mathfrak{m}$-convex. The intersection of any number of $\mathfrak{g}$-convex sets is again a $\mathfrak{g}$-convex set.

LEMMA 5.1. *A $\mathfrak{g}$-convex set $\mathfrak{a}$ is $\mathfrak{h}$-convex for any line $\mathfrak{h}$ disjoint with $\mathfrak{a}$.*

Proof. If p and q are different points of $\mathfrak{a}$, let $t = pq \cap \mathfrak{g}$, and let $u = pq \cap \mathfrak{h}$. Then u is an exterior point of $(pq)_t$ by assumption, hence $(pq)_u = (pq)_t$ is contained in $\mathfrak{a}$.

THEOREM 5.2. *Convex open sets define a topology of the plane of points, in which they are a base of open sets.*

Proof. Since any point of the plane is contained in some convex open set, it suffices to prove that for any two convex open sets $\mathfrak{a}$ and $\mathfrak{b}$ with a common point u, there exists a convex open set containing u and contained in $\mathfrak{a} \cap \mathfrak{b}$.

Let $\mathfrak{a}$ be $\mathfrak{l}$-convex, and let $\mathfrak{b}$ be $\mathfrak{m}$-convex, where $\mathfrak{l}$ and $\mathfrak{m}$ are two lines. If $\mathfrak{l} = \mathfrak{m}$, then $\mathfrak{a} \cap \mathfrak{b}$ is $\mathfrak{l}$-convex and also open, since the intersection of two open sets on a line is open. If $\mathfrak{l} \neq \mathfrak{m}$, let $\mathfrak{c}$ be the sector with sides $\mathfrak{l}$ and $\mathfrak{m}$

containing u. Then $\mathfrak{c}$ is $\mathfrak{l}$-convex and $\mathfrak{m}$-convex and open. Hence $\mathfrak{a} \cap \mathfrak{c}$ is $\mathfrak{l}$-convex and open, and also $\mathfrak{m}$-convex by Lemma 5.1. Thus $\mathfrak{a} \cap \mathfrak{c} \cap \mathfrak{b}$ is $\mathfrak{m}$-convex and open. It contains u and is contained in $\mathfrak{a} \cap \mathfrak{b}$.

The topology defined in Theorem 5.2 is called the *order topology* of the plane of points. The order topology of the plane of lines is defined dually. From now on, the plane of points and the plane of lines are considered as topological spaces with their order topologies. In the plane of points, lines are closed sets, and the relative topology on a line is its interval topology.

6. Convex quadrangles.

Definition. A *convex quadrangle* is the intersection of two sectors such that the vertex of each is an exterior point of the other.

If r and s are the vertices of two such sectors, then $r \neq s$, and both sectors are (rs)-convex open sets; hence so is their intersection. A convex quadrangle is thus an open convex set.

The intersection of two sectors with different vertices r and s is a convex quadrangle if and only if the line rs joining the two vertices is a common exterior line of the two angles corresponding to the sectors.

If a, b, c, d are four points, no three of which are collinear, let

$$r = ab \cap cd, \qquad s = ad \cap bc, \qquad t = ac \cap bd$$

be the diagonal points of the complete quadrangle $abcd$, and let

$$\rho_s = (ab\,; cd)_s, \qquad \sigma_r = (ad\,; bc)_r, \qquad \tau_r = (ac\,; bd)_r,$$
$$\rho_t = (ab\,; cd)_t, \qquad \sigma_t = (ad\,; bc)_t, \qquad \tau_s = (ac\,; bd)_s.$$

These notations will be used consistently.

If $\mathfrak{l} = ab$ and $\mathfrak{m} = cd$, then [3] $\mathfrak{lm} \,\|\, st$. Hence ρ_s and ρ_t are the two sectors with sides ab and cd, and $\rho_s \cap \sigma_r$ is a convex quadrangle containing t. It is easily seen that any convex quadrangle can be obtained in this fashion. This shows also that a convex quadrangle is never empty. The points a, b, c, d are called the *vertices*, the segments $(ab)_r$, $(bc)_s$, $(cd)_r$, $(ad)_s$, the *sides* of the quadrangle $\rho_s \cap \sigma_r$.

THEOREM 6.1. *The three convex quadrangles*

$$\rho_s \cap \sigma_r, \qquad \rho_t \cap \tau_r, \qquad \sigma_t \cap \tau_s,$$

their sides

$$(ab)_r, \quad (cd)_r, \quad (ad)_s, \quad (bc)_s, \quad (ac)_t, \quad (bd)_t,$$

and their vertices a, b, c, d, form a covering of the plane by disjoint convex sets.

[3] Cf. [1], section 3.21, p. 24.

Proof. Let $p = ab \cap st$. Then $ab \parallel pr$, and the segment $(ab)_r$ contains p and is contained in σ_r and in τ_r; hence it does not intersect one of the quadrangles. It is equally easily verified that all the sets of the covering are pairwise disjoint.

An exterior point of $(ab)_r$ on ab lies in $(ab)_p$, hence in $\sigma_t \cap \tau_s$. Thus a point on a side of the complete quadrangle $abcd$ always is in one of the sets of the covering.

Now let x be a point not lying on any side of the complete quadrangle $abcd$. Then x is in one of the sectors ρ_s and ρ_t, and we may assume $x \varepsilon \rho_s$. If x is in σ_r, x is in the quadrangle $\rho_s \cap \sigma_r$. Otherwise x is in σ_t. Let then $u = sx \cap ab$ and $v = sx \cap cd$. Then u is in $(ab)_p$ and hence in τ_s, and similarly v is in τ_s. Since $(uv)_s$ is the intersection of sx with ρ_s, x is in $(uv)_s$. But $(uv)_s$ is contained in the (st)-convex set τ_s, and so x is in the quadrangle $\sigma_t \cap \tau_s$. This completes the proof.

LEMMA 6.2. *The boundary of a convex quadrangle consists of its sides and vertices.*

Proof. The complement of the open set $\rho_t \cup \sigma_t$ consists of the quadrangle $\rho_s \cap \sigma_r$ with its sides and vertices, hence this is a closed set. If u is on a side of $\rho_s \cap \sigma_r$ or one of its vertices, let $v = ut \cap rs$. Then the segment $(ut)_v$ is contained in $\rho_s \cap \sigma_r$, and u is a boundary point of $(ut)_v$, hence also of $\rho_s \cap \sigma_r$.

LEMMA 6.3. *If p is in the segment $(ab)_r$, and q in $(cd)_r$, then one of the two segments with endpoints p and q lies in $\rho_s \cap \sigma_r$, and the other in $\rho_t \cap \tau_r$.*

Proof. Let $u = pq \cap rs$ and $v = pq \cap rt$. Then $pq \parallel uv$, and $(pq)_u$ is contained in $\rho_s \cap \sigma_r$, $(pq)_v$ in $\rho_t \cap \tau_r$.

If the dual of a convex quadrangle is called a *convex quadrilateral,* then the lines intersecting two segments on two different lines form a convex quadrilateral if and only if the point of intersection of the lines is a common exterior point for the segments. Thus the lines joining a point of $(ab)_r$ and a point of $(cd)_r$ form a convex quadrilateral. By Lemma 6.3, the lines of this quadrilateral do not meet the quadrangle $\sigma_t \cap \tau_s$ or its boundary. It is easily seen that every convex quadrilateral can be obtained in this fashion.

7. Continuity theorems.

LEMMA 7.1. *Let x, y, z be three points not on a line, and let α be an open convex set containing x. Then there is a complete quadrangle $abcd$ with diagonal points $r = y$ and $s = z$ such that the convex quadrangle $\rho_s \cap \sigma_r$ contains x and is contained in α together with its boundary.*

Proof. If $\mathfrak{a}$ is $\mathfrak{l}$-convex for a line $\mathfrak{l} \neq yz$, let $\mathfrak{b}$ be the sector with sides $\mathfrak{l}$ and yz containing x. Then $\mathfrak{a} \cap \mathfrak{b}$ is (yz)-convex and open. Hence we may assume $\mathfrak{a}$ to be (yz)-convex. Then there is a segment $(uv)_s$ on xz containing x and contained in $\mathfrak{a}$ together with its endpoints. Since the projection of uy on vy from z maps segments on segments, there are segments $(ab)_y$ on uy and $(cd)_y$ on vy, contained in $\mathfrak{a}$ together with their endpoints, such that $ab \parallel uy$ and $ad \cap bc = z$. Then $cd \parallel vy$, and $abcd$ is the desired quadrangle. For $r = y$, $s = z$, and x in $(uv)_s$ lies in $\rho_s \cap \sigma_r$. The vertices of $\rho_s \cap \sigma_r$ lie in the (rs)-convex set $\mathfrak{a}$; hence the quadrangle and its sides are contained in $\mathfrak{a}$.

Lemma 7.1 shows that the plane of points is a regular topological space, and that convex quadrangles form a base of open sets.

THEOREM 7.2. *The point of intersection of two different lines* $\mathfrak{l}$ *and* $\mathfrak{m}$ *is a continuous function of the pair* $(\mathfrak{l}, \mathfrak{m})$.

Proof. By Lemma 7.1 we may assume that a neighborhood of the point $\mathfrak{l} \cap \mathfrak{m}$ is a convex quadrangle $\rho_s \cap \sigma_r$ with r on $\mathfrak{m}$ and s on $\mathfrak{l}$. Then the lines intersectiong $(ab)_r$ and $(cd)_r$ form a neighborhood L of $\mathfrak{l}$, and the lines intersecting $(ad)_s$ and $(bc)_s$ form a neighborhood M of $\mathfrak{m}$. A line in L does not meet $\sigma_t \cap \tau_s$ or its boundary by Lemma 6.3, and a line in M does not meet $\rho_t \cap \tau_r$ or its boundary. Then it follows from Theorem 6.1 that the point of intersection of a line in L and a line in M must be in $\rho_s \cap \sigma_r$, and this proves the theorem.

THEOREM 7.3. *The line joining two different points* a *and* b *is a continuous function of the pair* (a, b).

This is the dual of Theorem 7.2.

8. Homeomorphism theorems. We shall use closed neighborhoods in this section rather than open neighborhoods. The closure of a convex quadrangle, consisting of the quadrangle with its sides and vertices, shall be called a *projective square.* The closure of $\rho_s \cap \sigma_r$ will be denoted by T, the closure of $\rho_t \cap \tau_r$ by S, and the closure of $\sigma_t \cap \tau_s$ by R. The *sides* of T are the intervals $[ab]_r$, $[bc]_s$, $[cd]_r$, and $[ad]_s$.

Any two intervals in the plane are homeomorphic, since projectivities are homeomorphisms. If the theorem of Desargues is valid in an ordered plane, then all intervals are homeomorphic to the unit interval in the ordered coordinate field or skew-field. If we speak of a homeomorphism between two intervals, we always assume that endpoints are mapped on endpoints.

THEOREM 8. 1. *A projective square is homeomorphic to the cartesian product of two intervals.*

Proof. For any point u in T, let $x = ab \cap us$ and $y = ad \cap ur$. Then x in $[ab]_r$ and y in $[ad]_s$ are continuous functions of u, and $u = xs \cap yr$. Thus u is also a continuous function of the pair (x, y), so that this correspondence is a homeomorphism between T and the product $[ab]_r \times [ad]_s$.

If $abcd$ and $a'b'c'd'$ are complete quadrangles in two ordered projective planes with homeomorphic intervals, then there is a homeomorphism between the boundaries of the projective squares T and T', by which the vertices and sides of T are mapped on the corresponding vertices and sides of T'. Such a homeomorphism shall be called a *p-homeomorphism,* where p means "proper." A homeomorphism between T and T' will be called a *P-homeomorphism* if it induces a p-homeomorphism between the boundaries.

LEMMA 8. 2. *There is a P-homeomorphism between T and T'.*

Proof. There are homeomorphisms ϕ between $[ab]_r$ and $[a'b']_{r'}$ and ψ between $[ad]_s$ and $[a'd']_{s'}$ such that $\phi(a) = \psi(a) = a'$. For any point u of T, let $x' = \phi(ab \cap us)$, $y' = \psi(ad \cap ur)$, and $\Phi(u) = x's' \cap y'r'$. Then Φ is a P-homeomorphism.

LEMMA 8. 3. *For every p-homeomorphism ϕ between the boundaries of T and of T', there is a P-homeomorphism Φ between T and T' that agrees with ϕ on the boundary.*

Proof. Since the product of two P-homeomorphisms is a P-homeomorphism, it suffices by Lemma 8. 2 to prove Lemma 8. 3 for $T = T'$. Again, it is sufficient to consider p-homeomorphisms which leave three sides of T fixed, since every p-homeomorphism on the boundary of T is a product of p-homeomorphisms of this special type.

Now let ϕ be a homeomorphism of $[ab]_r$ onto itself with $\phi(a) = a$ and $\phi(b) = b$. Let A be the closure of the open set $\tau_r \cap (ab;tr)_s$. Then A is contained in T, and its intersection with the boundary of T is the interval $[ab]_r$. The boundary of A consists of the three intervals $[ab]_r$, $[at]_c$, and $[bt]_d$. For u in A, $u \neq t$, the point $x = ab \cap tu$ is in $[ab]_r$. Now define $\Phi(u) = \phi(x)t \cap ru$ for u in A, $u \neq t$, and define $\Phi(u) = u$ for u in $T - A$ or $u = t$. Then $\Phi(x) = \phi(x)$ for x in $[ab]_r$, and $\Phi(u) = u$ for u in $[at]_c$ or in $[bt]_d$. It follows that Φ is continuous everywhere in T, except possibly at t.

If a'' is a point of $[at]_c$, then $b'' = bt \cap ra''$ is in $[bt]_c$, and b'',

$d'' = dt \cap sa''$, and $c'' = b''s \cap d''r$ are continuous functions of a''. If $a'' = t$, then $b'' = c'' = d'' = t$. Now let α be a convex neighborhood of t. We can choose a point a'' in $[at]_c$, different from t, such that a'', b'', c'', and d'' are in α. Then the projective square T'' with these vertices is a closed neighborhood of t contained in α. It follows from the construction that $\Phi(T'') = T''$, hence Φ is continuous at t.

Since the inverse mapping of Φ can be constructed in the same way as Φ, Φ is a P-homeomorphism of T onto itself which agrees with ϕ on $[ab]_r$ and leaves the other three sides of T fixed. This proves the lemma.

THEOREM 8.4. *Two ordered projective planes with homeomorphic intervals are homeomorphic.*

Proof. Let $abcd$ be a complete quadrangle in one plane, $a'b'c'd'$ a complete quadrangle in the other one. Then there is a mapping ϕ of the sides and vertices of R, S, and T on the sides and vertices of R', S', and T', which determines a p-homeomorphism for each of the three pairs of corresponding projective squares. Then by Lemma 8.3 there are homeomorphisms between R and R', between S and S', and between T and T' which agree with ϕ on the boundaries. Since two squares in the same plane have no common interior point, these homeomorphisms define a homeomorphism between the two planes.

COROLLARY 1. *In an ordered projective plane, the plane of lines is homeomorphic to the plane of points.*

COROLLARY 2. *If an interval in an ordered projective plane π is homeomorphic to the unit interval in an ordered field or skew-field K, then π is homeomorphic to the coordinate plane over the field K.*

It should be remarked that the homeomorphism of Theorem 8.4 need not be a collineation. In fact, section **9** gives an example of two ordered projective planes which are homeomorphic, but not collinear.

9. A nondesarguesian ordered plane. Let K be any ordered field or skew-field. We define points by homogeneous coordinates (x, y, z) over K with right multiplication, lines by homogeneous coordinates (a, b, c) with left multiplication.

A point (x, y, z) shall be on a line (a, b, c) with $ab \geqq 0$ if and only if

$$ax + by + cz = 0.$$

For a line (a, b, c) with $ab < 0$, we distinguish three cases. A point (x, y, z) shall be on this line if and only if

$$ax + by + cz = 0 \text{ for } xz \leqq 0,$$

$$\tfrac{1}{2}ax + by + cz = 0 \text{ for } x \text{ between } 0 \text{ and } z,$$

$$ax + by + (c - \tfrac{1}{2}a)z = 0 \text{ for } z \text{ between } 0 \text{ and } x.$$

This is a modification of the well known example of a nondesarguesian affine plane given by Hilbert and originally by F. R. Moulton.[4]. The verification of the axioms of incidence is quite easy, but rather lengthy. If we define a separation relation for pairs of points in the obvious manner, then the axioms S. 1 to S. 5 are readily verified.

Consider now the triangles abc and $a'b'c'$ with vertices

$$a: (0, 1, 2); \qquad b: (1, 1, 2); \qquad c: (1, 1, 1);$$

$$a': (0, -1, 2); \qquad b': (1, -1, 2); \qquad c': (1, -1, 1).$$

These triangles are perspective with center $(0, 1, 0)$, but the points

$$ab \cap a'b': (1, 0, 0); \qquad bc \cap b'c': (0, 0, 1); \qquad ac \cap a'c': (4, 1, -6)$$

are not collinear.

By Theorem 8. 4, Corollary 2, our plane is homeomorphic to the coordinate plane over the field K, but the two planes cannot be collinear.

NORTHWESTERN UNIVERSITY.

REFERENCES.

[1] H. S. M. Coxeter, *The real projective plane*, New York, 1949.

[2] ———, "Self-dual configurations and regular graphs," *Bulletin of the American Mathematical Society*, vol. 56 (1950), pp. 413-455.

[3] Marshall Hall, "Projective planes," *Transactions of the American Mathematical Society*, vol. 54 (1943), pp. 229-277.

[4] David Hilbert, *Grundlagen der Geometrie*, 7. Aufl., Leipzig und Berlin, 1930.

[5] F. R. Moulton, "A simple non-desarguesian plane geometry," *Transactions of the American Mathematical Society*, vol. 3 (1902), pp. 192-195.

[6] Ernst Steinitz, *Vorlesungen über die Theorie der Polyeder*, Berlin, 1934.

[7] Oswald Veblen and J. W. Young, *Projective geometry*, Boston, vol. I, 1910, vol. II, 1916.

[4] [4], p. 85, and [5].

MEANS IN GROUPS.*

By W. R. Scott.

1. Introduction. A number of authors have given sets of postulates for the arithmetic mean of n real numbers. Several of these sets of postulates have been given in purely algebraic form (see [2], [3], [4], [5], [7], [8]), and some of these latter are suitable for generalization to groups. Only the set due to Schimmack [7] will be discussed.

Let G be a (not necessarily Abelian) group written additively. Let $x_i \in G$, $i = 1, \cdots, n$, and let $f_n(x_1, \cdots, x_n) \in G$. Schimmack's postulates are:

(1) $\quad f_n(h + x_1, \cdots, h + x_n) = h + f_n(x_1, \cdots, x_n)$ for all $h \in G$.

(2) $\quad f_n(-x_1, \cdots, -x_n) = -f_n(x_1, \cdots, x_n)$.

(3) $\quad f_n$ is a symmetric function of $x_1, \cdots, x_n$.

(4) $\quad f_{n+1}(f_n, \cdots, f_n, x_{n+1}) = f_{n+1}(x_1, \cdots, x_{n+1})$.

In (4) and throughout the paper f_n will be used for $f_n(x_1, \cdots, x_n)$ whenever no confusion will result. A sequence $\{f_n\}$, $n = 1, 2, \cdots$, of functions will be called a *sequence* for brevity. A sequence satisfying (1), (2), (3), (4) will be called a *mean* on G.

Schimmack [7] showed that if G is the additive group R of reals, then the only mean is the ordinary arithmetic mean $f_n = (x_1 + \cdots + x_n)/n$. Beetle [1] showed that if $G = R$, then the postulates (1), (2), (3), (4) are completely independent. It will be shown here that, more generally, if G is an infinite Abelian torsion-free group such that $mG = G$ for all positive integers m, then Schimmack's and Beetle's conclusions hold. The question of existence and uniqueness of means, together with the complete independence of (1), (2), (3), (4), will be treated as completely as we are able.

2. Existence of means. If $\{f_n\}$ satisfies (1) and (2), then it follows readily that for all $h \in G$,

$$(1') \qquad f_n(x_1 + h, \cdots, x_n + h) = f_n(x_1, \cdots, x_n) + h.$$

If $\{f_n\}$ satisfies (4), then by induction, for $1 \leqq r \leqq n$,

$$(4') \qquad f_{n+1}(f_r, \cdots, f_r, x_{r+1}, \cdots, x_{n+1}) = f_{n+1}(x_1, \cdots, x_{n+1}).$$

* Received September 5, 1951.

THEOREM 1. *A group G possesses a mean if and only if*

(i) *the equation $mx = g$ possesses a unique solution (denoted by $x = g/m$) for every $g \varepsilon G$ and every positive integer m; and*

(ii) $y + (-y + nz)/(n+1) = z + (-z + ny)/(n+1)$ *for every $y \varepsilon G$, $z \varepsilon G$, $n \geqq 2$.*

If a mean exists, it is unique and is given by

$$\left\{\begin{array}{l} f_1(x_1) = x_1, \\ f_n = f_{n-1} + (-f_{n-1} + x_n)/n. \end{array}\right. \tag{5}$$

Proof. Parts of the argument are the same as in [7] but they will be given here for the sake of completeness.

Suppose first that a mean $\{f_n\}$ exists.

G is torsion-free. For if $g \neq 0$ is of finite order m, then by (1) and (3) we have

$$\begin{aligned} f_m(g, 2g, \cdots, (m-1)g, 0) &= g + f_m(0, g, \cdots, (m-1)g) \\ &= g + f_m(g, 2g, \cdots, (m-1)g, 0), \end{aligned}$$

which is a contradiction.

By (2), $f_n(0, \cdots, 0) = -f_n(0, \cdots, 0)$, whence $f_n(0, \cdots, 0) = 0$, since G is torsion-free. Hence $f_n(g, \cdots, g) = g$ by (1). Thus in particular $f_1(x_1) = x_1$, as asserted in (5).

We assert that

$$n f_n(g, 0, \cdots, 0) = g. \tag{6}$$

In fact, $f_2(g, 0) = g + f_2(0, -g) = g - f_2(0, g) = g - f_2(g, 0)$, and (6) is true for $n = 2$. Assume that (6) holds for $n = 2^s$. Let $f_{2^s}(g, 0, \cdots, 0) = y$. Thus $2^s y = g$. Let z be such that $2z = y$ (such a z exists by (6) with $n = 2$). Thus $2^{s+1} z = g$. Then by (4′)

$$\begin{aligned} &2^{s+1} f_{2^{s+1}}(g, 0, \cdots, 0) \\ &= 2^{s+1} f_{2^{s+1}}(f_{2^s}(g, 0, \cdots, 0), \cdots, f_{2^s}(g, 0, \cdots, 0), 0, \cdots, 0) \\ &= 2^{s+1} f_{2^{s+1}}(y, \cdots, y, 0, \cdots, 0) \\ &= 2^{s+1}(z + f_{2^{s+1}}(z, \cdots, z, -z, \cdots, -z)) = g, \end{aligned}$$

the last equality following from the fact that

$$f_{2^{s+1}}(z, \cdots, z, -z, \cdots, -z) = -f_{2^{s+1}}(z, \cdots, z, -z, \cdots, -z),$$

whence $f_{2^s+1}(z,\cdots,z,-z,\cdots,-z)=0$. Thus by induction (6) is true for $n=2^r$, $r=0,1,2,\cdots$. Suppose (6) true for $n+1$. Let $y=f_n(g,0,\cdots,0)$. Then

$$\begin{aligned} g &= (n+1)f_{n+1}(g,0,\cdots,0) \\ &= (n+1)f_{n+1}(f_n(g,0,\cdots,0),\cdots,f_n(g,0,\cdots,0),0) \\ &= (n+1)f_{n+1}(y,\cdots,y,0) = (n+1)(y+f_{n+1}(0,\cdots,0,-y)) \\ &= (n+1)y-y=ny, \end{aligned}$$

the next to the last equality holding since y and $f_{n+1}(0,\cdots,0,-y)$ commute, because $-y=(n+1)f_{n+1}(0,\cdots,0,-y)$. Hence (6) is true by induction.

It follows from (6) that the equation $mx=g$ has at least one root. For $m=1$ it has exactly one root. Again $f_2(2g,0)=g+f_2(g,-g)=g$, and (i) is true for $m=2$. Assume that (i) is true for $m=1,\cdots,n-1$, where $n\geqq 3$. Then by (1)-(4), (6),

$$\begin{aligned} &f_n(ny,0,\cdots,0) \\ &\quad = f_n(f_{n-1}(ny,0,\cdots,0),\cdots,f_{n-1}(ny,0,\cdots,0),0) \\ &\quad = f_n(ny/(n-1),\cdots,ny/(n-1),0) \\ &\quad = f_n(f_{n-1}(y,(n-1)y/(n-2),\cdots,(n-1)y/(n-2)), \\ &\qquad\qquad \cdots, f_{n-1}(y,(n-1)y/(n-2),\cdots,(n-1)y/(n-2)),0) \\ &\quad = f_n(y,(n-1)y/(n-2),\cdots,(n-1)y/(n-2),0) \\ &\quad = f_n(f_{n-1}((n-1)y/(n-2),\cdots,(n-1)y/(n-2),0), \\ &\qquad\qquad \cdots, f_{n-1}((n-1)y/(n-2),\cdots,(n-1)y/(n-2),0),y) \\ &\quad = f_n(y,\cdots,y)=y. \end{aligned}$$

Thus $ny=nz$ implies

$$y=f_n(ny,0,\cdots,0)=f_n(nz,0,\cdots,0)=z,$$

and (i) has been proved by induction.

For $n>1$,

$$\begin{aligned} f_n &= f_n(f_{n-1},\cdots,f_{n-1},x_n) \\ &= f_{n-1}+f_n(0,\cdots,0,-f_{n-1}+x_n) = f_{n-1}+(-f_{n-1}+x_n)/n, \end{aligned}$$

and (5) has been verified. The uniqueness of the mean follows from (i) and (5).

By (3), we have

$$f_{n+1}(x_1,\cdots,x_{n+1}) = f_{n+1}(x_1,\cdots,x_{n-1},x_{n+1},x_n), \tag{7}$$

whence by (5),

$$(8)\quad f_{n-1}+(-f_{n-1}+x_n)/n+((-x_n+f_{n-1})/n-f_{n-1}+x_{n+1})/(n+1)$$
$$=f_{n-1}+(-f_{n-1}+x_{n+1})/n+((-x_{n+1}+f_{n-1})/n-f_{n-1}+x_n)/(n+1).$$

Let y and z be given. Choose any $x_1,\cdots,x_{n-1}$, and then choose x_n and x_{n+1} so that

$$(9)\quad \begin{cases} ny=-f_{n-1}+x_{n+1} \\ nz=-f_{n-1}+x_n. \end{cases}$$

Making the substitutions (9) in equation (8), we get (ii).

Conversely suppose that (i) and (ii) are satisfied by G. Define $\{f_n\}$ by (5). Evidently $f_1(x_1)=x_1$ satisfies (1), (2), (3), (and (4) vacuously) for $n=1$.

Assume that (1) holds for a certain n. Then

$$f_{n+1}(h+x_1,\cdots,h+x_{n+1})$$
$$=f_n(h+x_1,\cdots,h+x_n)+(-f_n(h+x_1,\cdots,h+x_n)+h+x_{n+1})/(n+1)$$
$$=h+f_n+(-f_n-h+h+x_{n+1})/(n+1)=h+f_{n+1},$$

and (1) holds for all n.

Assume that (2) holds for given n. Then

$$f_{n+1}(-x_1,\cdots,-x_{n+1})$$
$$=f_n(-x_1,\cdots,-x_n)+(-f_n(-x_1,\cdots,-x_n)-x_{n+1})/(n+1)$$
$$=-f_n+(f_n-x_{n+1})/(n+1).$$

Now $(n+1)(-f_n+(f_n-x_{n+1})/(n+1)+f_n)=-x_{n+1}+f_n$. Hence

$$-f_n+(f_n-x_{n+1})/(n+1)=(-x_{n+1}+f_n)/(n+1)-f_n.$$

Thus

$$f_{n+1}(-x_1,\cdots,-x_{n+1})=(-x_{n+1}+f_n)/(n+1)-f_n$$
$$=-(f_n+(-f_n+x_{n+1})/(n+1))=-f_{n+1}.$$

Hence (2) holds for all n.

Re (4), it follows by induction that $f_n(x,\cdots,x)=x$. Therefore

$$f_{n+1}(f_n,\cdots,f_n,x_{n+1})$$
$$=f_n(f_n,\cdots,f_n)+(-f_n(f_n,\cdots,f_n)+x_{n+1})/(n+1)$$
$$=f_n+(-f_n+x_{n+1})/(n+1)=f_{n+1}.$$

Hence (4) holds for all n.

As noted above, (3) holds for $n = 1$. Now $-x_2 + x_1 + \frac{1}{2}(-x_1 + x_2) = \frac{1}{2}(-x_2 + x_1)$. Hence $x_1 + \frac{1}{2}(-x_1 + x_2) = x_2 + \frac{1}{2}(-x_2 + x_1)$, or $f_2(x_1, x_2) = f_2(x_2, x_1)$, and (3) holds for $n = 2$. Assume that (3) holds for a given $n \geqq 2$. Then to prove (3) inductively it is sufficient by (4) to prove (7), i. e. (8). But (8) follows from (ii) by means of the substitution (9). Thus (3) is proved by induction, and (5) defines a mean on G.

COROLLARY. *If G is an Abelian torsion-free group such that $mG = G$ for all integers $m > 0$, then G has a unique mean given by*

$$f_n(x_1, \cdots, x_n) = (x_1 + \cdots + x_n)/n.$$

Proof. It is easily verified that G satisfies conditions (i) and (ii) of Theorem 1. Equation (5) yields the above mean.

There remains the question of the existence of a non-Abelian group satisfying conditions (i) and (ii) of Theorem 1.

3. Complete independence. In order to discuss complete independence of the postulates (1)-(4) it is necessary to discuss the existence of 16 types of sequences. For brevity these types will be numbered as follows:

1. $++++$	5. $+-++$	9. $-+++$	13. $--++$
2. $+++-$	6. $+-+-$	10. $-++-$	14. $--+-$
3. $++-+$	7. $+--+$	11. $-+-+$	15. $---+$
4. $++--$	8. $+---$	12. $-+--$	16. $----$,

where a sequence is of type 7, for example, if (1) and (4) are true while (2) and (3) are false. The history (i. e. as to existence or non-existence) of each type of sequence will be discussed as completely as we are able. This has already been done for type 1: For the group G of order 1, clearly sequences of types 2 through 16 do not exist. *From now on assume that G has at least 2 elements.* Let $g \in G$, $g \neq 0$, be a fixed element.

LEMMA 1. *Every group G possesses sequences of types* 3, 4, 9, 10, 11, 12.

Proof. They will be exhibited.[1]

3. Let $f_n = x_1$.

4. Let $f_n = x_1$ if $n \neq 2$, and let $f_2 = x_2$. Re (4),

$$f_3(g, 0, 0) = g \neq 0 = f_3(f_2(g, 0), f_2(g, 0), 0).$$

[1] Several of the examples used in the proofs of Lemmas 1 and 2 were obtained by L. A. Colquitt and the author while working on the real case.

9. Let $f_n = 0$.

10. Let $f_2(0, x) = f_2(x, 0) = x$ for all $x \in G$, and let $f_n = 0$ otherwise. Concerning (4), we have $f_2(g, 0) = g \neq 0 = f_2(f_1(g), 0)$.

11. Let $f_1 = 0$ and $f_n = x_n$ for $n > 1$.

12. Let $f_1 = 0$ and $f_n = x_1$ for $n > 1$.

LEMMA 2. *A group G possesses sequences of types 7, 8, 13, 14, 15, 16 if and only if G has an element y of order greater than 2.*

Proof. If G has no element of order greater than 2, then $-x = x$ for all $x \in G$, and

$$f_n(-x_1, \cdots, -x_n) = f_n(x_1, \cdots, x_n) = -f_n(x_1, \cdots, x_n).$$

Therefore (2) is satisfied, and the listed types (and 5 and 6) do not exist.

Conversely suppose that G has an element y of order greater than 2. Then the following sequences will show the required existence.

7. Let $f_n = x_n + y$.

8. Let $f_n = x_1 + y$.

13. Let $f_n = y$.

14. Let $f_2(0, x) = f_2(x, 0) = x$ for all $x \in G$, and let $f_n = y$ otherwise. Re (4), we have $f_2(0, 0) = 0 \neq y = f_2(f_1(0), 0)$.

15. Let $f_n = -x_n + y$.

16. Let $f_n = -x_1 + y$.

LEMMA 3. *A sequence of type 6 exists if and only if G is torsion-free.*

Proof. Suppose that a sequence of type 6 exists. Then the first part of the proof of Theorem 1 shows that G must be torsion-free.

Conversely suppose that G is torsion-free. Let

$$(x_1, \cdots, x_n) \underset{h}{\sim} (y_1, \cdots, y_n)$$

if $x_r = h + y_{i_r}$, $r = 1, \cdots, n$, for some permutation $(i_1, \cdots, i_n)$ of $(1, \cdots, n)$. If also $(x_1, \cdots, x_n) \underset{h'}{\sim} (y_1, \cdots, y_n)$, then $h + y_{j_1} = h' + y_{j_2}$, $h + y_{j_2} = h' + y_{j_3}, \cdots, h + y_{j_m} = h' + y_{j_1}$. Hence $y_{j_1} = m(-h + h') + y_{j_1}$, and $m(-h + h') = 0$. But since G is torsion-free, $-h + h' = 0$, and $h = h'$.

Let $(x_1, \cdots, x_n) \sim (y_1, \cdots, y_n)$ if $(x_1, \cdots, x_n) \underset{h}{\sim} (y_1, \cdots, y_n)$ for

some (unique) h. Clearly $\sim$ is an equivalence relation. Choose a representative $(x_1^*, \cdots, x_n^*)$ in each equivalence class $\{(x_1, \cdots, x_n)\}$, choosing (g), $(g, 0)$ and $(g, -g)$ as the representatives of their respective equivalence classes, where g is a fixed element $\neq 0$. Note that $\{(g,0)\} \neq \{(g, -g)\}$. Let $(x_1, \cdots, x_n)$ be given. Then $(x_1, \cdots, x_n) \underset{h}{\sim} (x_1^*, \cdots, x_n^*)$. Define $f_n(x_1, \cdots, x_n) = h$. Now (1) and (3) are obviously satisfied. Re (2), $f_1(-0) = -g \neq g = -f_1(0)$. Re (4),

$$f_2(2g, 0) = g \neq 0 = f_2(g, 0) = f_2(f_1(2g), 0).$$

The only types not yet discussed are types 2 and 5. The results on these are only partial.

LEMMA 4. *If G possesses a sequence of type 2, then G is torsion free and possesses a unique solution x of the equation $2x = g$ for all $g \in G$.*

Proof. This was shown in the proof of Theorem 1, where (4) was not used to prove the existence and uniqueness of such solutions.

LEMMA 5. *Let G be such that*

(i) *$2x = g$ always has a solution for x;*

(ii) *$nx = g$, $n > 1$, has at most one solution for x for any $g \in G$. Then a sequence of type 2 exists.*

Proof. Because of its length, only an outline of the proof will be given.

Let $(x_1, \cdots, x_n) R_1 (y_1, \cdots, y_n)$ if there exist $h, k \in G$ and a permutation $(i_1, \cdots, i_n)$ of $(1, \cdots, n)$ such that

$$x_r = h + y_{i_r} + k, \qquad r = 1, \cdots, n.$$

Let $(x_1, \cdots, x_n) R_2 (y_1, \cdots, y_n)$ if there exist $h, k \in G$ and a permutation $(i_1, \cdots, i_n)$ of $(1, \cdots, n)$ such that

$$x_r = h - y_{i_r} + k, \qquad r = 1, \cdots, n.$$

Let $(x_1, \cdots, x_n) \sim (y_1, \cdots, y_n)$ if either $(x_1, \cdots, x_n) R_1 (y_1, \cdots, y_n)$ or $(x_1, \cdots, x_n) R_2 (y_1, \cdots, y_n)$. It follows that $\sim$ is an equivalence relation. In each equivalence class choose a representative $(x_1^*, \cdots, x_n^*)$, letting $(0, g)$, $(0, g, 3g)$, and $(\frac{1}{2}g, \frac{1}{2}g, 3g)$ be the representatives of their (distinct) equivalence classes, where g is a fixed element $\neq 0$.

Define f_n as follows:

$$f_n(x_1^*, \cdots, x_n^*) = \begin{cases} x_1^* + \frac{1}{2}(-x_1^* + x_{i_1}^*) \\ \qquad\qquad \text{if } (x_1^*, \cdots, x_n^*) R_2 (x_1^*, \cdots, x_n^*), \\ x_1^* \text{ otherwise;} \end{cases}$$

$$f_n(h + x_{j_1}^* + k, \cdots, h + x_{j_n}^* + k) = h + f_n(x_1^*, \cdots, x_n^*) + k,$$

$$f_n(h - x_{j_1}^* + k, \cdots, h - x_{j_n}^* + k) = h - f_n(x_1^*, \cdots, x_n^*) + k.$$

It can then be shown that f_n is well defined, and satisfies (1), (2) and (3). Re (4), we have

$$f_3(f_2(0, g), f_2(0, g), 3g) = f_3(\tfrac{1}{2}g, \tfrac{1}{2}g, 3g) = \tfrac{1}{2}g \neq 0 = f_3(0, g, 3g).$$

LEMMA 6. *If G can be ordered so that $x < y$ implies $h + x < h + y$ for all $h \varepsilon G$, then a sequence of type 5 exists.*

Proof. Define $f_n(x_1, \cdots, x_n) = \min(x_1, \cdots, x_n)$. Then (1), (3) and (4) are satisfied. Now $f_2(g, -g) = f_2(-g, g) = g$ or $-g$. Since $g \neq -g$ (such groups G are torsion-free), (2) is not satisfied.

See [6] for a discussion of such groups with the additional restriction that $x < y$ imply $x + h < y + h$ for all $h \varepsilon G$.

THEOREM 2. *If G is an Abelian torsion-free group such that $nG = G$ for all integers $n > 0$, then postulates* (1), (2), (3), (4) *are completely independent.*

Proof. Such a group G can be ordered (see [6]). Hence by Lemma 6 a sequence of type 5 exists. The other types exist by the Corollary to Theorem 1 and Lemas 1, 2, 3 and 5.

Sequences of types 2 and 6 can be replaced by simpler sequences in this case by making use of an ordering of G.

2. Let $f_n(x_1, \cdots, x_n) = \frac{1}{2}(\min + \max)$.

6. Let $f_n = \min + g$, where g is a fixed element $\neq 0$.

THEOREM 3. *If G is the additive group of reals, then there exists a unique mean on G, namely $f_n = (x_1 + \cdots + x_n)/n$, and* (1), (2), (3), (4) *are completely independent.*

This follows from Theorem 2 and the Corollary to Theorem 1. A similar theorem for the geometric mean of positive real numbers can be given, of course.

UNIVERSITY OF KANSAS.

REFERENCES.

[1] R. D. Beetle, "On the complete independence of Schimmack's postulates for the arithmetic mean," *Mathematische Annalen*, vol. 76 (1915), pp. ~~444-446~~.

[2] ———, *Bulletin of the American Mathematical Society*, vol. 22 (1916), pp. 276-277.

[3] E. L. Dodd, "The complete independence of certain properties of means," *Annals of Mathematics*, vol. 35 (1934), pp. 740-747.

[4] E. V. Huntington, "Sets of independent postulates for the arithmetic mean, the geometric mean, the harmonic mean, and the root mean square," *Transactions of the American Mathematical Society*, vol. 29 (1927), pp. 1-22.

[5] S. Narumi, "Note on the law of the arithmetical mean," *Tôhoku Mathematical Journal*, vol. 30 (1929), pp. 19-21.

[6] B. H. Neumann, "On ordered groups," *American Journal of Mathematics*, vol. 71 (1949), pp. 1-18.

[7] R. Schimmack, "Der Satz vom arithmetischen Mittel in axiomatischer Begründung," *Mathematische Annalen*, vol. 68 (1909), pp. 125-132.

[8] O. Suto, "Law of the arithmetical mean," *Tôhoku Mathematical Journal*, vol. 6 (1914), pp. 79-81

A PROOF OF THE MAXIMAL CHAIN THEOREM.*

By Orrin Frink.

The maximal chain theorem was first proved by Hausdorff in 1914 in [3] using transfinite induction. It states that every chain in a partially ordered set is contained in a maximal chain. It is equivalent to Zorn's lemma (*cf.* [1], [2], [5], [7], [8], [9], [10], [11], [13]). It was stated and proved long before Zorn's lemma, is somewhat simpler in form, and often just as convenient to use. Like Zorn's lemma, it owes its great usefulness to the fact that it allows one to avoid the theory of ordinal numbers and well-ordered sets in giving proofs in abstract mathematics. For this reason it would be desirable to have a proof of the theorem which is independent of the notion of a well-ordered set, and dependent only on the axiom of choice. However, all the proofs in the literature, either of the maximal chain theorem or of Zorn's lemma, seem to involve the notion of a well-ordering (*cf.* [1], [2], [5], [6], [7], [8], [9], [11]). The following proof does not involve the notion of a well-ordered set. It was suggested by Zermelo's second proof of the well-ordering theorem in [12].

Theorem. *Every chain of a partially ordered set is contained in a maximal chain.*

Proof. A chain is a simply ordered set; that is, of any two distinct elements of a chain, one necessarily precedes the other. Suppose the theorem false. Then in some partially ordered set P there is a chain A not contained in any maximal chain. Then corresponding to each chain C which includes A as a subset select, by means of the axiom of choice, a larger chain C', called the successor of C, containing only a single element of P not in C. This is possible since by assumption no chain which includes A is maximal.

We shall call a collection K of chains of P all of which include A *complete* if A is in K, the successor of each member of K is in K, and K contains the union of each chain of its chains. Clearly the collection of *all* chains which include A is complete, and the intersection of any set of complete collections is complete. Let J be the intersection of *all* complete collections of chains of P. Then J is the smallest complete collection. We wish to prove that J is a chain, which will involve a contradiction.

* Received July 11, 1951.

We shall call a chain C which is a member of the collection J *normal* if for every chain X of J either $X \subset C$ or $C \subset X$. It will be proved that every member of J is normal. If C is any normal member of J, define $K(C)$ to consist of all members X of J such that either $X \subset C$ or $C' \subset X$. Now the collection $K(C)$ is complete, since in the first place A is in $K(C)$ since $A \subset C$. Secondly, if X is in $K(C)$ so is its successor X'. For by the definition of $K(C)$, either $C' \subset X$ or $X \subset C$. If $C' \subset X$, then $C' \subset X'$. On the other hand, if $X \subset C$, then either $X' \subset C$ or $C \subset X'$, since C is normal. If $X' \subset C$, then X' is in $K(C)$. But if $C \subset X'$, then $X \subset C \subset X'$, and since X' has only a single element of P not in X, it follows that $C = X$ or $C = X'$, whence $C' \subset X'$ or $X' \subset C$. In either case X' is in $K(C)$, which therefore contains the successor of each of its members. Likewise $K(C)$ clearly contains the union of each chain of its chains, since the defining property $X \subset C$ or $C' \subset X$ of the collection $K(C)$ goes over to unions. It follows that $K(C)$ is complete, and since it is a subset of J, the smallest complete collection, $K(C)$ must be identical with J.

But by the defining property of the collection $K(C)$, it follows that the successor C' of every normal member C of J is also normal. Since the union of a chain of normal chains is clearly normal, it is seen that the collection of all normal members of J is complete, and hence is identical with the collection J itself, since J is the smallest complete collection. Hence J is a chain of chains, since all of its members are normal. Since J is complete and also a chain, it must contain the union U of all of its members, and likewise it must contain the successor U' of U, which is a proper superset of U. But this is impossible. This contradiction proves the theorem.

Conclusion. Zorn's lemma, which states that every collection of sets which contains the union of each chain of its sets has a maximal element, is an immediate consequence of the maximal chain theorem (*cf.* [1], [8], [9], [13]). Another formulation of Zorn's lemma states that if every chain of a partially ordered set P has an upper bound in P, then P contains a maximal element (*cf.* [1], [8], [9]).

The well-ordering theorem and the axiom of choice may also be derived as consequences of the maximal chain theorem by defining properly a partial ordering between the well-ordered subsets of a given set, or between the selection functions defined on subsets of a given set.

THE PENNSYLVANIA STATE COLLEGE.

REFERENCES.

[1] Garrett Birkhoff, *Lattice Theory*, revised edition, Colloquium Publications 25, New York, 1948, pp. 42-44.

[2] N. Bourbaki, *Éléments des mathématiques, Théorie des ensembles* I_1, Paris, 1939, pp. 36-37.

[3] F. Hausdorff, *Grundzüge der Mengenlehre*, Leipzig, 1914, pp. 140-141.

[4] G. Hessenberg, "Kettentheorie und Wohlordnung," *Journal für die reine und angewandte Mathematik*, vol. 135 (1908), pp. 81-133.

[5] H. Kneser, "Eine direkte Ableitung des Zornschen Lemmas aus dem Auswahlaxiom," *Mathematische Zeitschrift*, vol. 53 (1950), pp. 110-113.

[6] C. Kuratowski, "Une méthode d'élimination des nombres transfinis des raisonnements mathématiques," *Fundamenta Mathematica*, vol. 3 (1922), pp. 76-108.

[7] R. L. Moore, *Foundations of Point Set Theory*, Colloquium Publications 13, New York, 1932, pp. 84-85.

[8] Szele, T., "On Zorn's lemma," *Publicationes Mathematicae Debrecen*, vol. 1 (1950), pp. 254-257.

[9] J. W. Tukey, *Convergence and Uniformity in Topology*, Princeton, 1940, pp. 7-8.

[10] A. D. Wallace, "A substitute for the axiom of choice," *Bulletin of the American Mathematical Society*, vol. 50 (1944), pp. 278.

[11] E. Witt, "On Zorn's theorem," *Revista Matematica Hispano-Americana*, vol. 10, pp. 82-85 (1950).

[12] E. Zermelo, "Neuer Beweis für die Möglichkeit einer Wohlordnung," *Mathematische Annalen*, vol. 65 (1908), pp. 107-128.

[13] Max Zorn, "A remark on method in transfinite algebra," *Bulletin of the American Mathematical Society*, vol. 41 (1935), pp. 667-670.

NOTES ON LEFT DIVISION SYSTEMS WITH LEFT UNIT.*

By M. F. SMILEY.

1. R. Baer introduced in [3] the notion of *left division system with left unit* [2] by showing that these systems arise in a natural way from a simple method of multiplying the cosets of a subgroup of a group. It is our contention that many of the properties of loops [3] are valid also for these much more general systems. This is the first of a series of notes in which we shall support this contention. Herein we are concerned with basic structure theory culminating in the lemma of Zassenhaus. We owe the brevity [4] of the proofs in part to suggestions of a referee.

Let G be a system. We shall use $H \subseteq_s G$ as an abbreviation for "H is a subsystem of G." We observe that (Z1) $[e] \subseteq_s G$ *if e is the left unit of G*, and that (Z2) *the intersection of a family of subsystems of G is a subsystem of G*. We shall then denote the subsystem of G which is generated by H, $K \subseteq_s G$ by $\langle H, K\rangle$. If K is the kernel of a homomorphism η of G onto a system G', we shall write $K \subseteq_n G$. If $K \subseteq_n G$, we have (Z3) *if $H \subseteq_s G$, then $H\eta \subseteq_s G'$*, (Z4) *if $H' \subseteq_s G'$, then $H'\eta^{-1} \subseteq_s G$*, and (Z5) *if $H \subseteq_n G$, then $H\eta \subseteq_n G' = G\eta$*. In order to establish (Z5) we first prove that a subsystem K of G satisfies $K \subseteq_n G$ if and only if $(Kx)(Ky) = K(xy) = (Kx)y$ for every $x, y \varepsilon G$. (*Cf.* Baer [4], p. 455, Lemma 1.) When $K \subseteq_n G$, the mapping $x\eta \rightarrow Kx$ is an isomorphism of G', and $G/K = [Kx; x \varepsilon G]$.

Let us now mention a few immediate consequences of our definitions. If K is the kernel of a homomorphism η of G onto a system G', and $x\eta = y\eta$ for $x, y \varepsilon G$, then $x = ky$ for some $k \varepsilon K$. From (Z3) and (Z4) we see that $\langle H, K\rangle\eta = \langle H\eta, K\eta\rangle$ for $H, K \subseteq_s G$. We note that $K \subseteq_n G$ and $K \subseteq H \subseteq_s G$

* Received July 5, 1951; revised September 24, 1951.

[2] In the remainder of this note we shall use the word *system* in place of this longer phrase. Systems with unit were called *left loops* by Kiokemeister and Whitehead [13]. Their *admissible* left loops need not be normal in our sense, and there is no overlapping of our results for systems with theirs for left-loops.

[3] See [1, 2, 4-7, 10, 11].

[4] The use of the associative law will not shorten our proofs. It is, of course, well-known that an associative system is a group.

imply that $K \subseteq_n H$. The *modular law* $(KL \cap H) = (K \cap H)L$ for subsystems H, K, L of G with $L \subseteq H$ may be proved as in Baer [4].

The principal result of this note is the following lemma.

LEMMA. *If G is a system, $L \subseteq_n H \subseteq_s G$, and $K \subseteq_n G$, then (1) $H \cap K \subseteq_n H$, (2) $KL \subseteq_n KH$, (3) $(KL \cap H) \subseteq_n H$, (4) the mapping $(KL \cap H)h \to (KL)h$ with $h \in H$ is an isomorphism of $H/(KL \cap H)$ and KH/KL, (5) if we have $H \subseteq_n G$, then $KH = HK \subseteq_n G$, and (6) if we have $K \cap H \subseteq L$, then $Lh \to (KL)h$ with $h \in H$ is an isomorphism of H/L and KH/KL.*

Proof. Let η be a homomorphism of G onto a system G' with kernel K. By (Z3), $H\eta$ is a subsystem of G'. Thus η induces a homomorphism of H onto the system $H\eta$ and the kernel of this induced homomorphism is $H \cap K$. This proves (1). Now set $W = \langle H, K \rangle$. Then we have $W\eta = \langle H\eta, K\eta \rangle = H\eta$, and $W \subseteq KH$, $W = KH \subseteq_s G$. If also $H \subseteq_n G$, then $W = KH = HK$ (cf. Baer [4], p. 452). By (Z5), $L\eta \subseteq_n H\eta = (KH)\eta$. Let ϕ be a homomorphism of $H\eta$ onto a system H'' with kernel $L\eta$. Then $\eta\phi$ is a homomorphism of KH onto H'' with kernel KL. This proves (2). Again $\eta\phi$ induces a homomorphism of H onto H'' whose kernel is $KL \cap H$. This proves (3), which is also an immediate consequence of (1) and (2). If $h \in H$ and $k \in K$, we have $(kh)\eta\phi = h\eta\phi$, and it follows that $(KL)(kh) = (KL)h$, since $(hk)\eta\phi \to (KL)(kh)$ is an isomorphism of H'' and KH/KL. We then obtain (4) by noting that $h\eta\phi \to (KL \cap H)h$ is an isomorphism of H'' and $H/(KL \cap H)$. The statement (5) follows from (2) and our previous observation that $H \subseteq_n G$ implies $HK = KH$. Finally, (6) follows from (4) and the modular law.

COROLLARY (LEMMA OF ZASSENHAUS). *Let G be a system,*

$$A, B, A_1, B_1 \subseteq_s G, A_1 \subseteq_n A, B_1 \subseteq_n B.$$

Then

$$A_1(A \cap B_1) \subseteq_n A_1(A \cap B), \qquad B_1(B \cap A_1) \subseteq_n B_1(B \cap A),$$

and the identity mapping of G induces an isomorphism of the corresponding quotient systems.

Proof. Since $A \cap B \subseteq_s A$ and $A_1 \subseteq_n A$, (1) gives $A_1 \cap B \subseteq_n A \cap B$. Likewise, $B_1 \cap A \subseteq_n A \cap B$. Using (5), we see that

$$(A_1 \cap B)(A \cap B_1) = (A \cap B_1)(A_1 \cap B) = L \subseteq_n H = A \cap B.$$

We set $K = A_1$ and $G = A$, noting that $K \cap H = A_1 \cap B \subseteq L$, apply (6), and interchange A and B in the result.

Remarks. 1. If G is associative, then G is a group, and our discussion includes this case. On the other hand, if G is a loop, then $K \subseteq_n G$ does not imply that K is a normal subloop [5] in the sense of [1]. Thus our present exposition does not apply to loops.

2. However, it is possible to formulate a list of axioms which hold for systems and for loops and which justify our Lemma. We are indebted to R. Baer who suggested that such a list must exist. We consider a set $\mathfrak{J}$ of sets G with binary compositions which have a left unit e, for which $xa = b$ has a solution $x \varepsilon G$ for every $a, b \varepsilon G$, and such that $xa = a$ implies that $x = e$. If H is a subset of G which is an element of $\mathfrak{J}$ relative to the composition of G, we write $H \subseteq_t G$. Let $\mathfrak{Z}$ be a subset of $\mathfrak{J}$ such that the requirements (Z1)-(Z5) of our second paragraph hold. We agree, of course, to replace "system" by "element of $\mathfrak{Z}$" and "H is a subsystem of G" by "$H \subseteq_t G$ and H is an element of $\mathfrak{Z}$." We add the requirement: (Z6) If μ is a homomorphism of $G \varepsilon \mathfrak{Z}$ onto $G' \varepsilon \mathfrak{Z}$ with kernel K, then $K(xy) \subseteq (Kx)(Ky)$ for every $x, y \varepsilon G$. The interested reader will be able easily to adapt our proofs to these axioms if he adds the hypothesis $K \subseteq_n G$ to the statement of the modular law.

3. It is interesting to observe that a recent theorem of R. C. Buck [12] on homomorphisms is valid for multiplicative elements of the set $\mathfrak{J}$ of Remark 2. But trivial examples show that both our lemma and Buck's theorem fail in general.

4. Similarity (isotropy) [3, 1, 2, 11] will not reduce the study of systems to the study of loops. For if $G = [1, a, b]$, with 1 a left unit, $a1 = b$, $b1 = a$, $aa = ab = 1$, $ba = b$, $bb = a$, then G is a system which is not isotropic to a loop. The referee has remarked that every system is isotropic to a left loop; but, of course, not every left loop is a loop.

THE STATE UNIVERSITY OF IOWA.

[5] For a simple modification of the example of Bates and Kiokemeister [9] shows that a loop may have a system which is not a loop as a homomorphic image.

REFERENCES.

[1] A. A. Albert, "Quasigroups I," *Transactions of the American Mathematical Society*, vol. 54 (1943), pp. 507-519.

[2] ———, "Quasigroups II," *ibid.*, vol. 55 (1944), pp. 401-419.

[3] R. Baer, "Nets and groups I," *ibid.*, vol. 46 (1939), pp. 110-141.

[4] ———, "The homomorphism theorems for loops," *American Journal of Mathematics*, vol. 67 (1945), pp. 450-460.

[5] ———, "Splitting endomorphisms," *Transactions of the American Mathematical Society*, vol. 61 (1947), pp. 508-516.

[6] ———, "Endomorphism rings of operator loops," *ibid.*, vol. 61 (1947), pp. 517-529.

[7] ———, "Direct decompositions," *ibid.*, vol. 62 (1947), pp. 62-97.

[8] ———, "Direct decompositions into infinitely many summands," *ibid.*, vol. 64 (1948), pp. 519-551.

[9] G. E. Bates and F. Kiokemeister, "A note on homomorphic mappings of quasigroups into multiplicative systems," *Bulletin of the American Mathematical Society*, vol. 54 (1948), pp. 1180-1185.

[10] B. Brown and N. H. McCoy, "Some theorems on groups with application to ring theory," *Transactions of the American Mathematical Society*, vol. 69 (1950), pp. 302-311.

[11] R. H. Bruck, "Contributions to the theory of loops," *ibid.*, vol. 60 (1946), pp. 245-354.

[12] R. C. Buck, "A factoring theorem for homomorphisms," *Proceedings of the American Mathematical Society*, vol. 2 (1951), pp. 135-137.

[13] F. Kiokemeister and G. W. Whitehead, "A coset theory for left loops," (Abstract), *Bulletin of the American Mathematical Society*, vol. 51 (1951), pp. 60-61.

A CHARACTERIZATION OF FINITE DIMENSIONAL CONVEX SETS.*

By E. G. Straus and F. A. Valentine.

Let S be a *closed connected* set contained in a finite dimensional subspace of a linear space, and let R_n be the linear subspace of minimal dimension which contains S. A *maximal* convex subset of S is, by definition, one which is not contained in a larger convex subset of S. It is our purpose to establish the following result.

Theorem 1. *The set S defined above, is convex if and only if each point $x \in S$ is contained in a unique maximal convex subset $K(x)$ of S of dimension greater than or equal to $n-1$.* (Note: Observe that no restrictions are placed on the maximal convex subsets of S of dimension less than $n-1$).

Definition 1. The property placed on S in Theorem 1 is called property A. The symbol $|K(x)|$ denotes the maximal $(n-1)$-dimensional volume obtained from the class of all $(n-1)$-dimensional plane projections of $K(x)$. ($|K(x)|$ may be finite or infinite).

Lemma 1. *If there exists a non-convex closed connected set S with property A, then there exists a non-convex continuum[1] S_0 with property A, such that*

$$\operatorname*{g.l.b.}_{x \in S_0} |K(x)| = d > 0.$$

Proof. Choose a point $x_1 \in S$, and let C_m be the solid sphere in R_n of radius m with center at x_1. Also let $F_m \equiv S \cdot C_m$, and denote the component of F_m which contains x_1 by T_m. Since $\sum_{m=1}^{\infty} T_m = S$, and since S is non-convex, there exists a fixed value m such that T_m is not convex. Consider now the non-convex set T_{m+1}. In a well-established manner (*cf.* [2], Ch. 7), we may determine a topological space each element of which corresponds to one and only one of the maximal convex subsets of T_{m+1} of dimension $\geqq n-1$. By this process of régarding maximal convex subsets of T_{m+1} as points in a new

* Received August 20, 1951.

[1] A continuum is a compact connected set.

space, it is clear that T_{m+1} is mapped into a *set* called T^*_{m+1}, and it is also easily verified that T_m is mapped into a non-trivial *subcontinuum* T^*_m of T^*_{m+1}. (It should be noted that although T^*_m will be closed, bounded and connected, T^*_{m+1} may not be closed). To each point $x^* \varepsilon T^*_m$, we can assign the positive valued function $f(x^*) = |x^*| = |K(x)|$, where $K(x)$ is the maximal convex subset of T_{m+1} corresponding to x^*. This function $f(x^*)$ is upper semicontinuous on T^*_m. Hence the set $S^*_k = \{x^* \mid f(x^*) \geqq 1/k, x^* \varepsilon T^*_m\}$ is closed. If S^*_k were nowhere dense in T^*_m for each k, then $T^*_m = \sum_{k=1}^{\infty} S^*_k$ would be of the first category in itself, which contradicts the fact that it is a non-trivial continuum. Hence, there exists a value k such that S^*_k contains a non-trivial subcontinuum S^*_0. The pre-image S_0 of S^*_0 must then be a non-convex continuum in T_{m+1}, since S_0 is a closed connected subset of the compact set T_{m+1}. Moreover, S_0 satisfies property A, since each point $x \varepsilon S_0$ lies in a unique maximal convex subset $K(x)$ of T_{m+1} of dimension $\geqq n-1$ which lies in $S_0(K(x)$ must intersect $T_m)$. Also for each $x \varepsilon S_0$, we have $|K(x)| \geqq 1/k$, so that $d = \underset{x \varepsilon S_0}{\text{g.l.b.}} |K(x)| \geqq 1/k > 0$.

LEMMA 2. *If S_0 and d are the quantities defined in Lemma 1, then*

$$\underset{x \varepsilon S_0,\, \dim K(x) = n-1}{\text{g.l.b.}} |K(x)| = d.$$

Proof. The sets $K(x)$ with dimension $n-1$ are everywhere dense in S^*_0 since property A implies there exists at most a denumerable number of $K(x)$ with dimension n. This combined with the upper semicontinuity of $|K(x)|$ proves the lemma.

LEMMA 3. *Each plane P through the centroid of a bounded convex set K of dimension m and volume V divides K into two convex sets whose volumes V_1 and V_2 satisfy the inequalities*

$$V_i \geqq c_m V \qquad (i = 1, 2),$$

where c_m is a constant depending solely on m and not on K or P.

This lemma was proved for $m = 2$ by Neumann in [1], and for arbitrary m by Green and Gustin [unpublished].

LEMMA 4. *Consider the set S_0 in Lemma 1, and let $K(x)$ be a maximal convex subset of dimension $n-1$ such that $|K(x)| < (1 + c_{n-1})d$, where c_{n-1} is defined in Lemma 3. Let T_r be the right circular solid cylinder of radius r, whose axis passes through the centroid of $K(x)$ and is perpendicular to the plane of $K(x)$.*

Then there exists a neighborhood U of $K(x)$ and a value of r such that for each point $y \in U \cdot S_0$, the set $K(y)$ intersects all the elements of T_r.

Proof. If the lemma were false, there would exist a sequence of convex sets $K(y_i)$ whose distance from $K(x)$ would approach zero, and such that the following would hold. The sequence $K(y_i)$ would contain a subsequence which would converge to a convex subset C of $K(x)$ which does not contain the centroid of $K(x)$ in its interior. Due to the upper semicontinuity of $|K(x)|$, we have $|c| \geqq d$. Hence, the above with Lemma 3 implies that $|K(x)| \geqq |c| + c_{n-1}d \geqq (1 + c_{n-1})d$, which contradicts our hypothesis.

Sufficiency proof for Theorem 1. Assume that S is not convex, so that Lemmas 1 to 4 hold. We use the notation in Lemma 4. According to a known theorem in topology ([2], p. 16), there exists a non-trivial component C of $S_0 \cdot U$ containing $K(x)$. Since S_0 is non-convex, C is non-convex. Choose a point $y_1 \in C$. Since $K(y_1) \cdot T_r$ is a convex set dividing T_r into two parts, let L_1 be *any* line segment parallel to the axis of T_r, and having its endpoints in $K(y_1) \cdot T_r$ and $K(x)$ respectively. Let L be the line containing L_1. For any point $y \in C$, Lemma 4 implies $K(y) \cdot L \neq 0$. Hence, since C is a connected closed bounded set, property A implies that $C \cdot L$ is a closed line segment. Hence $L_1 \subset C$. But this implies that the portion of T_r between $K(x)$ and $K(y_1)$ is in C. However, this contradicts the fact that $K(x)$ is of dimension $n - 1$. This completes the sufficiency. The necessity of property A is obvious.

An interesting corollary to Theorem 1 for sets in R_2, the plane, is the following.

COROLLARY 1. *If S is a closed connected set in R_2, each point of which belongs to a unique maximal linear element (line segment, half-line or line) of S, then S is a linear element.*

Concluding remarks. One may ask in what respects our theorem is the best possible. It is the best possible in at least the following respects.

If we remove the assumption $\dim K(x) \geqq n - 1$, then the theorem no longer holds, as shown by the circular cylinder $x_1^2 + x_2^2 = 1$ in E_n.

It might be conjectured that our affine theorem has a purely topological origin. Thus one might think that if we replace the phrase "unique maximal convex subset of dimension $\geqq n - 1$" by "unique maximal closed (in the point set sense) surface of dimension $n - 1$" then the conclusion would be "S is an $n - 1$ dimensional closed surface." However, P. Erdös and A. H.

Stone have communicated to us the following counterexample for the case $n = 2$. Consider a Cantor set in the interval $(0, 1)$ of the X-axis. On each of the points of this set we erect a line segment of unit length perpendicular to the X-axis on which $y \geqq 0$. Consider the intervals in the complement of the Cantor set. For each interval I, draw the diagonal segment joining the upper end point of the vertical segment at the left endpoint of I to the right endpoint of I. The resulting set S of vertical segments plus diagonal segments is connected, and every point in S belongs to a unique maximal closed arc (either a vertical segment or a polygonal line consisting of two vertical segments plus a diagonal segment). However, S is *not* an arc.

The following question is still unsettled. Suppose $S \subset R_n$ is a closed connected set such that each of its points is contained in a unique maximal closed connected $(n-1)$-dimensional subset of S which is contained in an $(n-1)$-dimensional hyperplane of R_n. Is then S an $(n-1)$-dimensional set which is contained in an $(n-1)$-dimensional hyperplane? If $n = 2$, the question is answered in the affirmative by Corollary 1. For $n > 2$, the question is still undecided.

UNIVERSITY OF CALIFORNIA AT LOS ANGELES.

REFERENCES.

[1] B. H. Neumann, "On an invariant of plane regions and mass distributions," *Journal of the London Mathematical Society*, vol. 20 (1945), pp. 226-237.

[2] G. T. Whyburn, *Analytic Topology*, American Mathematical Society Colloquium Publications (1942).

ON ADDITIVE IDEAL THEORY IN GENERAL RINGS.[1]

By Charles W. Curtis.

Introduction. It is the purpose of this paper to present some contributions to the structure theory of non-commutative ideal lattices, as developed by Krull [10][2] and Dilworth [3]. The results of Part 1, except for (5) of Lemma 1.1 and the first part of Theorem 1.4, hold in an arbitrary non-commutative residuated lattice in the sense of Dilworth [3]. The results are stated only for ideals[3] in rings, however, since our interest is the application of these results to the structure theory of rings. The main result of the paper is the determination in **1.4** of the maximal elements in the inclusion ordered set of right associated primes of an ideal. In **1.1**, **1.2** **1.3** a decomposition theory of ideals is worked out, similar to the work of Fuchs [5] for commutative rings. In **1.5** Krull's theory [10] of right associated prime ideals is linked to the theory of primary ideals. The notion of isolated component ideal leads to a new approach to the uniqueness theory of primary ideals. In **1.6** it is proved that if the ring R satisfies the ascending chain condition (A. C. C.) for right ideals, and if every ideal in R is a finite intersection of primary ideals, then the intersection of the powers of the Jacobson radical is the zero ideal. The methods presented here do not lead to a proof of this theorem for arbitrary rings with A. C. C. for right ideals because of an example due to E. Noether, published in [10], of a ring with A. C. C. for right ideals having the property that not every ideal is an intersection of primary ideals.

I wish to thank Professor N. Jacobson for the encouragement and many helpful suggestions he has given me during the preparation of this paper.

Part 1. The General Theory.

1.0. Notations. In this paper R will denote a non-commutative ring with a unit element 1. Proper ideals in R will be denoted by $A, B, C, \cdots$, and elements of R by $x, y, a, b, \cdots$. We shall use the symbol 0 both for

[1] Received June 29, 1951.

[2] Numbers in brackets refer to the list of references at the end of the paper.

[3] "ideal" always means "two-sided ideal."

the zero ideal and for the zero element of R. $\{\Sigma C \mid C$ has the property $P\}$ means the join of those ideals C having the property P, while $\{x \mid x$ has the property $P\}$ means the set of all elements x of R having the property P.

1.1. Primal ideals.

Definition 1.1. The *quotient* AB^{-1} of the ideals A and B is defined by $AB^{-1} = \{\Sigma C \mid CB \subseteq A\}$. Similarly $B^{-1}A = \{\Sigma C \mid BC \subseteq A\}$.

We observe that $AB^{-1} = \{x \mid xB \subseteq A\}$; $B^{-1}A = \{x \mid Bx \subseteq A\}$. The following properties of the quotients are well known ([10], [3]).

LEMMA 1.1.

(1) $(AB^{-1})B \subseteq A$, $AB^{-1} = R$ *if* $B \subseteq A$, *and if* $B_1 \subseteq B_2$ *then* $AB_1^{-1} \supseteq AB_2^{-1}$.

(2) $(AC^{-1})B^{-1} = A(BC)^{-1}$.

(3) $(A_1 \cap \cdots \cap A_n)B^{-1} = A_1B^{-1} \cap \cdots \cap A_nB^{-1}$,

(4) $A(B_1 + \cdots + B_n)^{-1} = AB_1^{-1} \cap \cdots \cap AB_n^{-1}$, *and more generally, if* $\{B_\mu\}$ *is an arbitrary collection of ideals,*

(5) $A(\Sigma_\mu B_\mu)^{-1} = \bigcap_\mu AB_\mu^{-1}$.

Similar rules hold for the quotients $B^{-1}A$.

Definition 1.2. If $AB^{-1} = A$ $(B^{-1}A = A)$ then B *is right* (left) *prime* to A.

Definition 1.3. A is *right primal*[4] if the join P of the ideals not right prime to A is again not prime to A. If A is primal, then P is called the *adjoint ideal* of A.

Let R be a commutative ring. If G is primal in the sense of Definition 1.3 then G is primal in the sense of Fuchs [5]. The converse is false.[5]

[4] Although there is also a theory of left primal ideals, we shall consider only right primal ideals. Henceforth, "primal" means "right primal," and "prime to A" means "right prime to A."

[5] This point is settled by the following example, due to the referee. Let R be the set of all finite expressions $a_1t^{n_1} + \cdots + a_rt^{n_r}$, with coefficients in a field, and exponents non-negative rational numbers. With respect to the obvious definitions of addition and multiplication, R is a commutative ring in which the principal ideal generated by t is primal in the sense of Fuchs, but not primal in the sense of Definition 1.3.

Definition 1.4. P is a *prime ideal* [6] in an arbitrary ring R, if $AB \subseteq P$ implies $A \subseteq P$ or $B \subseteq P$.

From Lemma 1.1, (2) we have

LEMMA 1.2. *If A is primal, then the adjoint ideal P of A is a prime ideal.*

Definition 1.5. A is *strongly irreducible* if A cannot be expressed as an intersection, finite or infinite, of proper divisors of A. A is *irreducible* if A cannot be expressed as an intersection of a finite number of proper divisors of A.

From Lemma 1.1, (5) and (4), we obtain at once

LEMMA 1.3. *Every strongly irreducible ideal is right primal. If R satisfies the A.C.C. for ideals, then every irreducible ideal is right primal.*

THEOREM 1.4. *Every ideal in R is the intersection of its primal divisors. If R satisfies the A.C.C. for ideals, then every ideal in R is an intersection of a finite number of primal ideals.*[7]

Proof. Let A be an ideal in R. Since there are no ideals not prime to the ring R itself, we shall agree that R is a primal ideal. In order to prove the first part of the theorem it is sufficient to prove that if $c \not\in A$, then there exists a primal divisor G of A not containing c. From Zorn's lemma, however, it follows that there exists a divisor G of A having the property that every proper divisor of G contains c. G is therefore strongly irreducible, and by Lemma 1.3, primal. The second part of the theorem is an immediate consequence of the A.C.C. and Lemma 1.3.

1.2. Uniqueness of primal decompositions.

Definition 1.6. (E. Noether [17]) The intersection $A = G_1 \cap \cdots \cap G_n$ is *irredundant* if no G_i divides its *complement*

$$G_1 \cap \cdots G_{i-1} \cap G_{i+1} \cap \cdots \cap G_n.$$

[6] This definition is due to Krull [10].

[7] The first right principal components defined by Krull in [10] are primal ideals, and therefore a theorem of [10], which states that every ideal is the intersection of its first right principal components, is an instance of Theorem 1.4. I am indebted to the referee for pointing out to me that since an ideal A is strongly irreducible if and only if the ring R/A is subdirectly irreducible in the sense of Birkhoff ("Subdirect unions in universal algebra," *Bulletin of the American Mathematical Society*, vol. 50 (1944), pp. 764-768), Theorem 1.4 is a consequence of Birkhoff's result, which states that an arbitrary ring is a subdirect sum of subdirectly irreducible rings.

The intersection is *reduced* if no G_i can be replaced by a proper divisor. Intersections which are both irredundant and reduced are called *normal.*

Throughout the remainder of this section we shall assume that R satisfies the A. C. C. for ideals. The idea of the next result is due to Fuchs [5].

THEOREM 1. 5. *Let $A = G_1 \cap \cdots \cap G_n$ be a reduced representation*[8] *of A by primal ideals G_i with adjoint prime ideals P_i. An ideal B is not prime to A if and only if B is contained in one of the P_i.*

Proof. By (4) we have

$$AB^{-1} = (G_1 \cap \cdots \cap G_n)B^{-1} = G_1B^{-1} \cap \cdots \cap G_nB^{-1}.$$

Since the intersection is reduced, $AB^{-1} \neq A$ if and only if $G_iB^{-1} \neq G_i$ for some i. But $G_iB^{-1} \neq G_i$ if and only if $B \subseteq P_i$, and the proof is complete.

THEOREM 1. 6. *The reduced intersection of a finite number of primal ideals $A = G_1 \cap \cdots \cap G_n$ with adjoint ideals $P_1, \cdots, P_n$ is primal if and only if one prime P_j divides all the others.*

Proof. First let some P_i, say P_1 divide all the other P_j. By Theorem 1. 5, if $AB^{-1} \neq A$, then $B \subseteq P_i \subseteq P_1$ for some i; hence if S denotes the join of the ideals not prime to A, $S \subseteq P_1$. But again by Theorem 1. 5 we conclude that $AS^{-1} \neq A$, and hence A is primal.

Conversely, let S be the adjoint ideal of A. Since $AS^{-1} \neq A$, $S \subseteq P_i$ for some i. But $\sum_{j=1}^{n} P_j \subseteq S \subseteq P_i$, and the theorem is proved.

THEOREM 1. 7. *Every ideal in R is a normal intersection of primal ideals, such that no adjoint prime divides another.*

Proof. By the A. C. C. A is an intersection of irreducible ideals, and we can assume that the intersection is irredundant. By Lemma II of E. Noether's paper [17] which holds, together with the other results we shall require from that paper, in the non-commutative case, the intersection is necessarily reduced. By Lemma 1. 3, the ideals appearing in the intersection are primal ideals. Let their adjoint primes be $P_1, \cdots, P_n$, and suppose the indices chosen so that $P_1, \cdots, P_k$ are the maximal elements in the set $\{P_i\}$, ordered by inclusion. If we replace the intersection G_1 of those primal ideals whose adjoint primes are P_1 or a multiple of P_1 by G_1 itself, then G_1 is, by Theorem 1. 6, a primal

[8] By a representation of A, we mean an expression of A as an intersection of some of its divisors.

ideal, and by Lemma IV of [17] the resulting intersection is still reduced. Next we replace the intersection G_2 of those ideals, not already incorporated into G_1, whose adjoint ideals are P_2 or a multiple of P_2 by G_2, and again by Lemma IV of [17] the intersection is reduced. By repeating this process k times, we obtain at last a normal representation of A by primal ideals having the desired properties.

THEOREM 1. 8. *If A has two normal representations by primal ideals, such that no adjoint prime in either representation divides another in the same representation, then the adjoint primes in the two representations and the number of components are the same.*

Proof. Let $A = G_1 \cap \cdots \cap G_n = G_1^* \cap \cdots \cap G_m^*$ be two normal primal representations of A satisfying the hypotheses of the theorem. Let the adjoint primes be $P_1, \cdots, P_n$ and $P_1^*, \cdots, P_m^*$. We shall prove that P_1 is contained in some P_j^*. Since both representations are reduced, we can apply Theorem 1. 5 once to conclude that P_1 is not prime to A, and again to conclude that P_1 is contained in some P_j^*. By symmetry we can show that $P_j^* \subseteq P_k$ for some k. Thus $P_1 \subseteq P_j^* \subseteq P_k$, which contradicts our hypothesis unless $P_1 = P_j^* = P_k$. The rest of the proof is now clear.

1. 3. Maximal prime ideals. McCoy defined a set S in a ring R to be an m-system if $a, b \in S$ imply the existence of an element $x \in R$ such that $axb \in S$. The empty set is considered to be an m-system. He then defined the *radical* $M(A)$ of an arbitrary ideal A in R to be the set of elements r such that every m-system containing r contains an element of A. McCoy proved ([15], Theorem 2) that the radical $M(A)$ of an ideal A is the intersection of the minimal prime divisors of A, thus achieving a successful generalization of the work of Krull [11]. Levitzki has sharpened this result by proving in [14] that $M(A)/A$ is actually the *lower radical* in the sense of Baer [2] of R/A.

For commutative rings, Fuchs has characterized in [6] the structure of the intersection of all maximal prime ideals belonging to an arbitrary ideal. For non-commutative rings R satisfying the A. C. C. for ideals, we shall prove a result analogous to Fuchs' theorem.

Definition 1. 7. Let P be a divisor of A, where $A \neq R$. P is a *maximal prime ideal* belonging to A if (i) $AP^{-1} \neq A$, and (ii) if Q is a proper divisor of P, then $AQ^{-1} = A$. P is a *minimal prime ideal* of A if P is a prime ideal, and if there exists no prime ideal Q such that $A \subseteq Q \subseteq P$.

It should be observed that any ideal satisfying (i) and (ii) is necessarily a prime ideal, and consequently the definition is meaningful in the form in which we have stated it.

In order to conclude that A has at least one maximal prime ideal, it is enough to assume the A. C. C. for ideals. *Throughout this section we shall assume the A. C. C. for ideals.*

Definition 1. 8. The join S of all ideals C such that $A(C+B)^{-1} \neq A$ whenever $AB^{-1} \neq A$ is called the *adjoint ideal* of A (compare Fuchs [6]).

Since $A(A+B)^{-1} = AA^{-1} \cap AB^{-1} = R \cap AB^{-1} = AB^{-1}$, we see that $A \subseteq S$. Until now we have spoken of the adjoint ideal only in connection with primal ideals. We shall prove that if A is a primal ideal with adjoint prime S', then $S = S'$. In fact if $c \in S$, then (c)[9] is certainly not prime to A; hence $(c) \subseteq S'$, and we conclude that $S \subseteq S'$. Conversely, let $AB^{-1} \neq A$. Then $A(S'+B)^{-1} = A(S')^{-1} \neq A$; hence $S' \subseteq S$.

THEOREM 1. 9. *The adjoint ideal of A is the intersection of all maximal prime ideals of A.*

Proof. Let c be contained in the adjoint ideal S of A, and suppose that a maximal prime ideal P of A does not contain c. Then $P + (c)$ is a proper divisor of P, and since $AP^{-1} \neq A$, $A(P+(c))^{-1} \neq A$, contrary to our assumption that P is a maximal prime belonging to A. Conversely, let d be contained in every maximal prime of A, and let B be any ideal not prime to A. Then by the A. C. C. B is contained in a maximal prime, say P^*, and $B + (d) \subseteq P^*$; hence $B + (d)$ is not prime to A. Thus (d) and hence d is contained in the adjoint ideal of A.

By virtue of a remark of McCoy ([15], page 829) any prime ideal containing A contains a minimal prime of A. Since the McCoy radical is the intersection of all the minimal primes of A, it follows from this remark and Theorem 1. 9 that the adjoint ideal of A always contains the McCoy radical.

We have defined the set of maximal primes ideals belonging to A independently of a particular representation of A by primal ideals. The next theorem shows the connection between maximal prime ideals and primal representations.

THEOREM 1. 10. *Let $A = G_1 \cap \cdots \cap G_n$ be a normal primal decomposition of A, with adjoint primes $P_1, \cdots, P_n$. Then an ideal P is a maximal prime ideal belonging to A if and only if P is one of the P_i.*

---

[9] (c) is the principal ideal RcR generated by c.

Proof. We shall prove first that the P_i are maximal primes belonging to A. By Theorem 1.5, $AP_i^{-1} \neq A$ for each i, and again by Theorem 1.5, since the representation is reduced, each divisor of P_i is prime to A, $1 \leq i \leq n$. Consequently the P_i are maximal primes belonging to A. Conversely if P is any maximal prime belonging to A, then $AP^{-1} \neq A$, and by Theorem 1.8, $P \subseteq P_i$ for some i. From the maximality of P and the fact that $AP_i \neq A$, we conclude that $P = P_i$.

COROLLARY 1.11. *A is primal if and only if A has exactly one maximal prime ideal.*

1.4. Associated prime ideals. In the additive ideal theory of commutative rings it is desirable to give a definition of the prime ideals "associated" with a given ideal, independently of the notion of primary ideal. Krull gave such a definition for commutative rings in [9], and for non-commutative rings in [10]. We shall follow his methods in this section.

Definition 1.9. The ideal I is a (right) *isolated component ideal* (I.C.I.) of A if there exists an ideal B and an integer $q > 0$ such that $I = AB^{-q} = AB^{-q-1} = \cdots$.

We shall assume throughout this section that R satisfies the A.C.C. for ideals. If A is a given ideal, and B an arbitrary ideal, then we have in general $AB^{-1} \subseteq AB^{-2} \subseteq \cdots$, and by the A.C.C. there exists an integer q such that $AB^{-q} = AB^{-q-1} = \cdots$; the I.C.I. AB^{-q} is called the I.C.I. *generated* by B, and we shall denote it by $I(A, B)$, or more simply, when it is clear from the context that A is the basic ideal in the discussion, by $I(B)$.

Definition 1.10. A prime ideal P is a (right) *associated prime ideal* of A if (i) $I(P)$ is a proper divisor of A, and (ii) $I(P)^{-1}A \subseteq P$.

LEMMA 1.12. *Every maximal prime ideal P of A is a right associated prime ideal of A.*

Proof. Since $AP^{-1} \neq A$, $I(P)$, which contains AP^{-1}, is a proper divisor of A. It remains to prove that $I(P)^{-1}A \subseteq P$, that is, if C is an ideal such that $I(P)C \subseteq A$ then $C \subseteq P$. Let $D = AP^{-1}$; then D is a proper divisor of A, and $DP \subseteq A$. Since $D \subseteq I(P)$, we have also $DC \subseteq A$. Combining these results we obtain $D(P + C) \subseteq A$, and consequently $A(P + C)^{-1} \neq A$. Since P is a maximal prime of A, $P + C \subseteq P$, and we have $C \subseteq P$.

LEMMA 1.13. *Every right associated prime of A is contained in a maximal prime of A.*

Proof. This result follows immediately from the fact that if P is a right associated prime of A, then $AP^{-1} \neq A$.

From Lemmas 1.12 and 1.13 we obtain at once

THEOREM 1.14. *A prime ideal P is a maximal prime ideal of A if and only if P is a maximal element in the inclusion-ordered set of right associated prime ideals of A.*

1.5. Primary ideals. In this section we assume that R satisfies the A. C. C. for ideals.

Definition 1.11. An ideal Q is (right) *primary* if for arbitrary ideals A and B, $AB \subseteq Q$, $A \not\subseteq Q$ implies $B^r \subseteq Q$ for some positive integer r. As usual "primary" means "right primary" from now on. If Q is primary then the ideal

$$P = \{\Sigma B \mid B^r \subseteq Q \text{ for some positive integer } r\}$$

is called *the prime ideal belonging to Q*, or simply *the prime ideal* of Q.

It is easy to verify that since R satisfies the A. C. C., the ideal P defined above actually is a prime ideal.

If Q is primary, then Q is primal with adjoint prime P. The converse is false, as an example given in [5], page 2 indicates.

Let A be an ideal which is a finite intersection

$$A = Q_1 \cap \cdots \cap Q_k \tag{6}$$

of primary ideals. Dilworth has observed ([3], Theorem 6.1) that the intersection (6) can be refined by uniting those ideals Q_i having the same prime, obtaining an expression for A as an irredundant intersection $Q_1^* \cap \cdots \cap Q_m^*$ of primary ideals Q_i^*, where $Q_i^* \cap Q_j^*$ is not primary if $i \neq j$. Such an intersection we shall call a *shortest representation of A by primary ideals*, or more briefly, a S. R of A. Dilworth's result also states that for any S. R. of A, the primes and the number of primary ideals are uniquely determined. These results have also been announced by Murdoch [16].

Let $A = Q_1 \cap \cdots \cap Q_k$ be a S. R. of A by primary ideals Q_i with primes P_i. Consider a subset S of the set $\{P_i\}$ having the property that if $P_i \in S$, then $P_j \subseteq P_i$ implies $P_j \in S$. The intersection of the primary ideals belonging to the primes in S is called a *Noether isolated component* of A.

The methods of Krull [9] lead directly to a proof of the next result.

THEOREM 1.15. *If B is an ideal, then $AB^{-1} = AB^{-2}$ if and only if*

$A_1 = AB^{-1}$ is the ring R or a Noether isolated component of A appearing in every S. R. of A. Furthermore, if A_1 is any Noether isolated component of A, there exists an ideal B such that $A_1 = AB^{-1}$.

COROLLARY 1.16. *A_1 is a right isolated component ideal of A if and only if A_1 is a Noether isolated component of A.*

The following theorem is an application of Theorem 1.15 and the methods of Krull [9]. Since the previous theorem can be proved independently of Theorem 6.1 of [3], Part (b) of the next result (which is an immediate corollary of Part (a)) gives an alternative approach to Theorem 6.1 of [3].

THEOREM 1.17.

(a) *The set of right associated primes of A is identical with the set of prime ideals belonging to the primary ideals in every S. R. of A.*

(b) *The number of primary ideals in a S. R. of A and the primes belonging to them are uniquely determined.*

(c) *The Noether isolated components of A are uniquely determined by their associated primes.*

1.6. On the powers of an ideal. Let R be a ring satisfying the A. C. C. for right ideals, and having the property that every ideal in R is an intersection of a finite number of primary ideals. If A is an ideal in R, let $A^\omega = \bigcap_{i=1}^{\infty} A^i$. The first part of the proof of Satz 1 of [12] can be transferred to a ring satisfying the above hypotheses, proving

LEMMA 1.18. $A^\omega A = A^\omega$.

THEOREM 1.19. *If J is the Jacobson radical of R then $J^\omega = 0$.*

Proof. Lemma 1.18 and a result of Jacobson [8, Theorem 10].

Part 2. On Some Examples.

2.1. We shall say that R has a *Noetherian ideal theory* if i) R satisfies the A. C. C., and ii) every ideal in R is an intersection of a finite number of (right) primary ideals. The following example, mentioned in the introduction, shows that not every ring with A. C. C. has a Noetherian ideal theory. Let K be the field of rational numbers, and let R be the algebra over K with

basis elements e_1, e_2, n and the multiplication table $e_i^2 = e_i$, $i = 1, 2$; $n^2 = 0$; $e_1e_2 = e_2e_1 = 0$; $e_1n = ne_2 = n$; $e_2n = ne_1 = 0$. A straightforward determination of the ideals in R leads to a verification of the fact that the zero ideal of R is neither primary nor an intersection of primary ideals.

In view of this example and the fact that the results of 1.5 and 1.6 are valid for rings with a Noetherian ideal theory, it is important to consider examples of rings having a Noetherian ideal theory. It follows easily from the fundamental homomorphism theorem that if R has a Noetherian ideal theory, so has any homomorphic image of R. Another class of examples is furnished by finite matrix rings over rings with a Noetherian ideal theory.

From the results of Fitting [4] it follows directly that if R is a ring satisfying the A. C. C. for ideals, and having the property that every right or left ideal is a two-sided ideal, then R has a Noetherian ideal theory. This class of rings contains all Noetherian rings.[10]

Let R be a ring with A. C. C. for ideals, satisfying the conditions that if P is a prime ideal in R, different from the zero ideal, then P is maximal and if P_1 and P_2 are prime ideals in R, then $P_1P_2 = P_2P_1$. We shall apply the theory of primal ideals to prove that R has a Noetherian ideal theory. It is sufficient to prove that if A is an ideal in R (possibly the zero ideal), then A is a finite intersection of primary ideals. Let $A = Q_1 \cap \cdots \cap Q_n$ be a representation of A by primal ideals Q_i with adjoint prime ideals P_i. We shall prove that each Q_i is primary. In fact, consider Q_1. Let $\bar{P}_1$ be a minimal prime ideal of Q_1. Since the prime ideals commute, it follows from [10], Theorem 6 that $Q_1\bar{P}_1^{-1} \neq Q_1$, and hence $\bar{P}_1 \subseteq P_1$. Either $\bar{P}_1 = 0$ and hence $A = 0$, which shows that A is prime, or $\bar{P}_1 = P_1$. In the latter case Q_1 has the unique maximal and minimal prime P_1, and it follows that Q_1 is primary. Similarly the ideals $Q_2, \cdots, Q_n$ are primary. Examples of rings of this type are non-commutative principal ideal rings, and more generally, orders satisfying the axioms of Asano.[11]

We shall digress for a moment to consider other types of ideals related to primary ideals. We say that Q is *strongly right primary* if $ab \in Q$, $a \notin Q$ imply $b^r \in Q$ for some positive integer r.

In the commutative case, this definition coincides with Definition III given by E. Noether in [17], while our Definition 1.11 coincides with Definition IIIa of E. Noether. For commutative rings the two definitions

---

[10] A Noetherian ring is a commutative ring with unit element, satisfying the A. C. C. for ideals.

[11] Cf. [7], Chapters III and VI.

are equivalent if R satisfies the A. C. C., but this is no longer true for non-commutative rings.

Let R satisfy the A. C. C. for right ideals. If Q is strongly primary, then Q is primary, and we shall give an example to show that the converse is false. Let Q be strongly primary, and suppose $AB \subseteq Q$, $A \not\subseteq Q$, where A, B are ideals in R. By definition B/Q is a nil ideal in the ring R/Q, which also satisfies the A. C. C. for right ideals. By a result of Levitzki [13] every nil ideal in R/Q is nilpotent, and this proves that Q is primary. For the example, let D be a finite dimensional central simple algebra, and let $R = D_n$, the ring of n by n matrices with coefficients in D, for $n > 1$. R is a simple algebra, and its zero ideal is prime, and a fortiori primary. It is easy to find zero divisors in R.

Murdoch [16] has defined an ideal Q to be primary if $aRb \subseteq Q$, $a \notin Q$ implies $b \in M(Q)$, where $M(Q)$ is the McCoy radical of Q. We shall prove that for rings satisfying the maximum condition for right ideals, Murdoch's definition is equivalent to Definition 1.11. In fact, let Q be primary according to 1.11, and let $aRb \subseteq Q$, where $a \notin Q$. It follows that the ideal RbR is nilpotent modulo Q; hence $RbR \subseteq M(Q)$, and $b \in M(Q)$, since $M(Q)$ is a radical ideal. Conversely let Q be primary in the sense of Murdoch, and let $AB \subseteq Q$, $A \not\subseteq Q$ where A and B are ideals. It follows that B is a nil ideal modulo Q, and by Levitzki's result again, B is nilpotent modulo Q.

2.2. Let S be a simple ring with unit element. Then it is well known that the center Φ of S is a field, and S is a central simple algebra over Φ. Let $S[X_1, \cdots, X_n]$, or more briefly $S[X]$, denote the algebra over Φ of polynomials in n indeterminates X_i, where we assume that the indeterminates are commutative with one another and with the elements of S. With the same assumptions concerning the X_i, let $S\{X_1, \cdots, X_n\}$, or more briefly $S\{X\}$, denote the algebra over Φ of formal power series in the variables X_i. Let R be either $S[X]$ or $S\{X\}$. In this section we shall prove that if either $R = S[X]$ and S is arbitrary, or if $R = S\{X\}$ and $(S:\Phi) < \infty$, then R has a Noetherian ideal theory.

Let M be the monomial basis for R, that is M consists of all monomials $X_1^{e_1}X_2^{e_2} \cdots X_n^{e_n}$ where the e_i are non-negative integers. Let $f \in R$. Then f can be written uniquely in the form $f = \Sigma a(m)m$, $a(m) \in S$, $m \in M$, where $a(m)$ is always a finitely valued function if $R = C[X]$, and $a(m)$ is an arbitrary function if $R = S\{X\}$. Let $B = (b, c, \cdots)$ be a basis for S over Φ. Then $a(m)$ can be written uniquely as $a(m) = \Sigma\lambda(m, b)b$, $\lambda(m, b) \in \Phi$, where the functions $\lambda(m, b)$ are b-finitely valued, that is the matrix $(\lambda(m, b))$ is row finite. We have

$$(1) \qquad f = \Sigma_m (\Sigma_b \lambda(m, b) b) m = \Sigma_{m,b} \lambda(m, b) bm$$

uniquely, and conversely if $(\mu(n, c))$ is any row finite matrix on $M \times B$ to Φ, $g = \Sigma\mu(n, c)cn$ is an element of R. Let $bc = \Sigma_d \xi(b, c, d)d$, and $mn = \Sigma_p \eta(m, n, p)p$ (for each b, c, $\xi(b, c, d)$ is d-finitely valued, and for each m, n, $\eta(m, n, p)$ is p — 1-valued.) Then

$$(2) \qquad fg = (\Sigma\lambda(m, b)bm)(\Sigma\mu(n, c)cn) = \sum_{(m,n,b,c,d,p)} \lambda(m, b)\mu(n, c)\xi(b, c, d)\eta(m, n, p)\,dp,$$

where the matrix

$$(\theta(p, d)) = \Big(\sum_{(m,n,b,c)} \lambda(m, b)\mu(n, c)\xi(b, c, d)\eta(m, n, p)\Big)$$

is row finite.

Now consider the vector space V over Φ of all mappings on the product set $B \times M \to \Phi$. If $b \otimes m$ is the mapping which assigns 1 to the pair (b, m) and zero to everything else, then every element x of V can be expressed uniquely in the form

$$(3) \qquad x = \Sigma\lambda(m, b) b \otimes m.$$

The scalar multiplication, of course, is given by

$$(4) \qquad \alpha x = \Sigma\alpha\lambda(m, b) b \otimes m, \qquad \alpha \varepsilon \Phi.$$

Consider the subspace W of V consisting of all vectors (3) for which $(\lambda(m, b))$ is a row finite matrix. If we define a multiplication in W by means of (2), then with respect to (2) and (4), W is an algebra over Φ.

Let us assume that some b is the unit element 1 of S. Then the subalgebra of W consisting of all vectors $\Sigma\lambda(1, m) 1 \otimes m$, where $\lambda(1, m)$ is always m-finitely valued or m-infinitely valued depending upon whether R is a polynomial or a power series algebra, is isomorphic to $\Phi[X]$ or $\Phi\{X\}$ respectively, and we shall denote it by $1 \otimes \Phi[X]$ (resp. $1 \otimes \Phi\{X\}$.) Similarly the set of vectors $\Sigma\lambda(b, 1) b \otimes 1$, where $1 = X_1^0 X_2^0 \cdots X_n^0$, and where $\lambda(b, 1)$ is always b-finitely valued, forms a subalgebra of W isomorphic to S, which we shall denote by $S \otimes 1$. The algebra W^* generated by *finite* sums of products of elements from $1 \otimes \Phi\{X\}$ and $S \otimes 1$ does not equal W unless S is finite dimensional, but in the polynomial case $W^* = W$ without any restriction on S.

In the polynomial case W is simply the Kronecker product of the algebras S and $\Phi[X]$, but in the power series case, when $(S : \Phi) < \infty$, W is some sort of a generalized Kronecker product. Let $D = \Phi[X]$ or $\Phi\{X\}$ depending

upon whether $R = S[X]$ or $S\{X\}$. Then we shall write $W = S \otimes D$, and call it the Kronecker product[12] of S and D.

From (1), (2), (3), and (4) we see that the mapping

$$\Sigma\lambda(b, m)bm \rightarrow \Sigma\lambda(b, m)b \otimes m \tag{5}$$

is an isomorphism of R onto $S \otimes D$. We require now the following result.

THEOREM. *If U is an ideal in D, then $S \otimes U$ is an ideal in $S \otimes D$, and the mapping $U \rightarrow S \otimes U$ is a lattice isomorphism of the lattice of ideals of D onto the lattice of ideals of $S \otimes D$, provided either $(S:\Phi) < \infty$ or $D = \Phi[X]$.*

In the polynomial case, where $S \otimes D$ is the usual Kronecker product, the result was proved by Nakayama and Azumaya [1]. When $(S:\Phi) < \infty$, the following observation and method used by Jacobson[13] in proving the theorem of Nakayama and Azumaya, leads to a proof for the power series case. The essential point in Jacobson's proof is that if $v \in S \otimes D$, then v can be expressed as a finite sum $v = \sum_{i=1}^{n} a_i \otimes f_i$, $a_i \in S$, $f_i \in D$, where the a_i are linearly independent in S. From our remarks above, this can be done if and only if $(S:\Phi) < \infty$.

If the conditions of the theorem hold, then it is easy to verify that if A and B are ideals in D, then

$$S \otimes AB = (S \otimes A)(S \otimes B). \tag{6}$$

From the Theorem and (6) it follows that if Q is a primary ideal in D, then $S \otimes Q$ is primary in $S \otimes D$. Since D is a Noetherian ring, the theorem impiles that $S \otimes D$ has a Noetherian ideal theory, and finally, from (5), since R is isomorphic to $S \otimes D$, we conclude that R has a Noetherian ideal theory.

UNIVERSITY OF WISCONSIN.

---

[12] It is not difficult to show that the structure of W is independent of the particular bases we have chosen for S and D.

[13] Cf. Theorem 3, Chapter VI, of a book on the structure theory of rings, to appear in the Annals of Mathematics Studies.

REFERENCES.

[1] G. Azumaya and T. Nakayama, "On irreducible rings," *Annals of Mathematics*, vol. 48 (1947), pp. 949-965.

[2] R. Baer, "Radical ideals," *American Journal of Mathematics*, vol. 65 (1943), pp. 537-568.

[3] R. Dilworth, "Non-commutative residuated lattices," *Transactions of the American Mathematical Society*, vol. 46 (1939), pp. 426-444.

[4] H. Fitting, "Primärkomponentenzerlegung in nichtkommutativen Ringe," *Mathematische Annalen*, vol. 111 (1935), pp. 19-41.

[5] L. Fuchs, "On primal ideals," *Proceedings of the American Mathematical Society*, vol. 1 (1950), pp. 1-8.

[6] L. Fuchs, "On a special property of the principal components of an ideal," *Det Kongelige Norske Videnskabers Selskab*, vol. XXII (1950), pp. 28-30.

[7] N. Jacobson, "*The theory of rings*," Mathematical Surveys, Number 2, 1943.

[8] ———, "The radical and semi-simplicity for arbitrary rings," *American Journal of Mathematics*, vol. 67 (1945), pp. 300-320.

[9] W. Krull, "Ein neuer Beweis für die Hauptsätze der allgemeinen Idealtheorie," *Mathematische Annalen*, vol. 90 (1923), pp. 55.

[10] ———, "Zweiseitige Ideale in nichtkommutativen Bereichen," *Mathematische Zeitschrift*, vol. 28 (1928), pp. 481-503.

[11] ———, "Idealtheorie in Ringen ohne Endlichkeitsbedingung," *Mathematische Annalen*, vol. 101 (1929), pp. 729-744.

[12] ———, "Dimensiontheorie in den Stellenringen," *Journal für die Reine und Angewandte Mathematik*, vol. 179 (1938), pp. 204-226.

[13] J. Levitzki, "On multiplicative systems," *Compositio Mathematica*, vol. 8 (1950), pp. 76-80.

[14] ———, "Prime ideals and the lower radical," to appear in the *American Journal of Mathematics*.

[15] N. H. McCoy, "Prime ideals in general rings," *American Journal of Mathematics*, vol. 71 (1948), pp. 823-833.

[16] D. Murdoch, "Intersections of primary ideals in a non-commutative ring," *Bulletin of the American Mathematical Society*, vol. 56 (1950), Abstract 456.

[17] E. Noether, "Idealtheorie in Ringbereichen," *Mathematische Annalen*, vol. 83 (1921), pp. 24-66.

TWO DECOMPOSITION THEOREMS FOR A CLASS OF FINITE ORIENTED GRAPHS.*

By R. Duncan Luce.

1. Introduction.[1] The object of study in this paper is the class of finite oriented graphs which are subject to the conditions:

i. at most two branches exist between any pair of nodes (vertices), and

ii. whenever two branches do exist between a pair of nodes they shall have the opposite orientation.

Such a system will be called a network. The justification for introducing this word is its wide use in those applied sciences where oriented graphs of this type are playing an important role; for example: electrical networks, sociometric networks or diagrams, abstract programs for digital computers, and the neural networks of mathematical biology.

It is convenient to give a self-contained definition: A *network* N of order m is a system composed of two sets M and P, M being a finite non-empty set of m elements called the *nodes* of N and P a prescribed subset of the set of all ordered pairs of nodes. The members of P (i. e. the oriented branches) are called the *links* of N. The number of links of a network N will be denoted by $p(N)$, or simply by p when no ambiguity can arise. To indicate that N is of order m and has p links we shall write $N = N(m, p)$. Lower case Latin letters such as $a, b, c, \cdots$ will be used for nodes, and bracketed ordered pairs $(ab), (ca), \cdots$ to denote links. If (ab) is a link, the first node, a, will be called the *initial node* and the second, b, the *end node* of the link.

* Received October 2, 1951.

[1] Several of the concepts defined in the introduction have been assigned terms by D. König, *Theorie der Endlichen und Unendlichen Graphen*, New York, Chelsea Publishing Co., 1950. A brief glossary with page references to König is presented:

node	= Knotenpunkt, p. 1
link	= Gerichtete Kante, p. 4
disjoint	= Fremd, p. 3
arc of a network	= Zweifache Kante, p. 93
link of the form (aa)	= Schlinge, p. 3
non-reflexive graph	= Graph im engeren Sinn, p. 4
circuit	= Zyklus, p. 29
chain which is not a circuit	= Bahn, p. 30.

A *subnetwork* N' of a network N is a subset M' of the nodes, M, of N, with P' taken to be some subset (not necessarily proper) of those links of N which are definable on M'. If $M' = M$, we shall say the subnetwork is *complete*. Two subnetworks of a given network are *disjoint* if they have no nodes, and therefore no links, in common.

Each network is obviously a binary relation over a finite set, its nodes, and conversely every binary relation over a finite set can be interpreted as a network. This allows us to present all examples as relation matrices with entries 0 and 1 from the two element Boolean algebra. Furthermore, this suggests that if N and N' are two networks over the same (or isomorphic) set of nodes M, then by $N - N'$ we shall mean the complete subnetwork of N having those links of N which are not links of N'. If N' is a subnetwork of N, and if N' has the set of links P', then by the network formed from N by the *removal* of the links P', we mean $N - N'$. If N' has but one link, (ab), of N, then we shall write $N - N' = N - (ab)$.

We shall call a network *non-reflexive* if there are no links of the form (aa).

In case both the links (ab) and (ba) are present in a network, we shall say that an *arc ab* exists between a and b, the arc consisting of this pair of links, each of which will be said to be a member of the arc. This terminology is justified by the fact that when every link is a member of an arc the network is isomorphic (in the obvious sense of the word) to a graph without 2-circuits, to use a term of Whitney [2]; this is what we shall mean by saying that *a network is a graph*. Observe that the arcs of a network N are not the same as the branches (or arcs) of the graph which is oriented to form N. A link of the form (aa) is always the arc aa.

A (connected and oriented) *q-chain* from a to b is a set of q links of the form $(ac_1), (c_1c_2), \cdots, (c_{q-2}c_{q-1}), (c_{q-1}b)$, such that no node appears more than once, except in the case $a = b$ where a appears twice. Any q-chain from a to b will be denoted by (ab, q). Observe that $(ab, 1) = (ab)$. If c is a node included in a q-chain from a to b, then we may subdivide the chain into the "product" of two chains, one from a to c, and the other from c to b, i. e., $(ab, q) = (ac, q')(cb, q - q')$, $q' < q$.

An (oriented) *circuit* is a chain of the form (aa, q). A circuit of two links is an arc and conversely.

A network is *connected* if there exists a chain from each node to every other node. A network which is not connected is *disconnected*. When N is

[2] Whitney, H., "Non-separable and planar graphs," *Transactions of the American Mathematical Society*, vol. 34 (1932), p. 339.

treated as an oriented graph, connectedness is defined topologically; our definition implies topological connectedness but is not implied by it. However, as applied to networks which are graphs, the two definitions are equivalent.

2. A decomposition theorem for arbitrary networks. In this section we shall give four related definitions for networks of order m which will depend on the integers k between 0 and m. These definitions will be used in their full generality in Theorem 2.4, which shows that, in a certain sense, we need only consider the definitions in the case $k = 1$. Consequently the rest of the paper will be, for the most part, devoted to that special case.

First we need a measure of how easily a connected network is disconnected by the removal of links. We shall say a network is of degree 0 if it is not connected. A network is of *degree* k, $1 \leq k \leq m$, if there exists a set of k distinct links whose removal from the network will result in a complete subnetwork of degree 0, while the removal of any set of $q < k$ links results in a complete connected subnetwork.[3] The degree of a network is unique.

LEMMA 2.1. *If $N(m, p)$ is a network of degree k, then $p \geq km$.*

Proof. It will obviously suffice to show that each node is the initial node of at least k links. This is true, for if not, then the removal of the links for which such a node is the initial node will disconnect N. This contradicts the assumption that the degree is k.

In addition to the concept of degree, we need a condition implying that there is an even distribution of connectedness throughout the network; roughly, that the degree of any connected subnetwork is not greater than that of the network itself. That this is not always the case is evidenced by any graph formed of an $m - 2$ simplex, $m \geq 4$, and a single node joined by a single arc to one of the nodes of the simplex. The network is of degree 1, and the simplex, which is a connected subnetwork, is of degree $m - 2 \geq 2$. A definition which will suffice is the following. A network is said to be *k-minimal,* $1 \leq k \leq m$, if the removal of any link results in a complete subnetwork of degree $k - 1$. The existence of such networks is proved in Lemma 2.3. A network is *k-uniform* if every connected subnetwork is of degree $\leq k$. If a network is 1-uniform and connected we say it is *uniform.*

LEMMA 2.2. *If N is a k-minimal network with $k \geq 2$, then N is k-uniform and of degree k.*

[3] This definition of degree has no relation to the *Grad* defined by König, *op. cit.*, p. 3.

Proof. Let S be any connected subnetwork of N, and suppose it has degree d. Let $(ab) \varepsilon S$. $N - (ab) = N'$ is of degree $k - 1$, so in N' there exists a set U of $k - 1 \geq 1$ links, whose removal from N' results in a complete disconnected subnetwork, N''. If in N'' there is a chain from a to b, we may replace (ab) and still have a disconnected network N^* which is formed from N by removing the links of U. In that case, the removal of any $(cd) \varepsilon U$ from N results in a complete subnetwork of degree $k - 2 \geq 0$, since $k \geq 2$. This is contrary to the assumption that N is k-minimal, so there is no chain from a to b. Thus, the removal of no more than k links, those of U which are in S and (ab), from S, implies a is not connected to b by any chain. It follows that $d \leq k$.

Specifically, N has degree $d \leq k$. If $d \leq k - 1$, then, since the removal of any link results in a complete subnetwork of degree $k - 1$, it follows that $d = k - 1$. Let U be a set of $k - 1$ links whose removal from N results in a complete disconnected subnetwork. U is non-empty since $k \geq 2$. Remove $(ab) \varepsilon U$ *from* N. The resulting network is, by definition, of degree $k - 1$; however, the remaining $k - 2$ links of U disconnect $N - (ab)$. Hence $d = k$.

We note that the above argument does not apply for $k = 1$; in fact, any disconnected network is 1-minimal, since the removal of any link results in a complete subnetwork of degree 0. But some networks are both connected and 1-minimal; these we shall call ***minimal***. A minimal network is clearly non-reflexive and uniform.

LEMMA 2. 3. *If N is a network of degree $k \geq 1$, then for every integer q, $1 \leq q \leq k$, there exists a complete connected subnetwork of N which is q-minimal.*

Proof. Let C_q be the set of all complete connected subnetworks of N having degree q. Since N is finite and $q \leq k$, it is obvious that C_q is non-empty. Let

$$p_q = \max_{N' \varepsilon C_q} p(N - N').$$

Since N is finite, there exists some $N_q \varepsilon C_q$ such that $p(N - N_q) = p_q$. N_q is, by choice, connected. Hence it will suffice to show that N_q is q-minimal. Suppose the removal of some link does not result in a complete subnetwork of degree $q - 1$. Then, since the removal of one link cannot lower the degree by more than 1, the resulting network N' has degree q. Thus $N' \varepsilon C_q$ and $p(N - N_q) < p(N - N') \leq p_q$, which is contrary to choice.

A complete connected subnetwork N' of N such that, in the terms of the

above proof, $p(N-N')=p_q$, is called a *q-descendant* of N. If N_q and N'_q are two q-descendants of a network N, then $p(N_q)=p(N'_q)$. It is clear that every connected network has at least one 1-descendant, but this is not generally true for $q>1$. Because of their importance the 1-descendants will be called simply *descendants*. It is clear that a descendant is minimal.

A network N will be called the sum of complete subnetworks N_i, $i=1,2,\cdots,t$, and written $N=\sum_{i=1}^{t} N_i$, if each link of N is contained in exactly one of the N_i.

THEOREM 2.4 (*first decomposition theorem*). *To every network N there exists a unique number k, its degree, and at least one set of $k+1$ complete 1-minimal subnetworks, N_i, such that*

i. $N=\sum_{i=1}^{k+1} N_i$,

ii. *N_{k+1} is disconnected,*

iii. *if $k\geq 1$, then N_1 is minimal,*

iv. *$\sum_{i=1}^{j} N_i$ is a j-descendant of $\sum_{i=1}^{j+1} N_i$, $1\leq j\leq k$,*

and

v. *the connected subnetworks of the N_i, $1\leq i\leq k$, are minimal, and so these networks N_i are 1-uniform.*

Proof. By definition there is a unique degree k assigned to every network. If $k=0$, then N is not connected and we are done. If $k>0$, select, according to Lemma 2.3, a k-descendant N'_k of N, and define $N_{k+1}=N-N'_k$. N_{k+1} is not connected; for if so, N is the sum of two complete subnetworks having, respectively, degree k (Lemma 2.2) and degree ≥ 1. This, we will show, implies that N has degree $\geq k+1$, which is contrary to assumption.

To show this we prove the slightly more general statement: If $N=N_1+N_2$, and these networks have degrees k, k_1, and k_2 respectively, then $k\geq k_1+k_2$. For, by definition, there exists a set U having k links, such that their removal from N results in a complete disconnected subnetwork N', and this is not true for any smaller set. Of these k links, let u_1 be in N_1, and u_2 in N_2. By the definition of a sum, $k=u_1+u_2$. Furthermore, $u_1\geq k_1$, since we may remove from N first the links of U and then the remaining links of N_2. This complete subnetwork, which obviously is N_1 minus u_1 links, is disconnected, so $u_1\geq k_1$. Similarly, $u_2\geq k_2$, whence the result.

In the network N'_k select a $(k-1)$-descendant, N'_{k-1}, and let $N_k=N'_k$

$- N'_{k-1}$. N_k is 1-minimal, for if $(ab) \varepsilon N_k$, let $N^*_k = N'_k - (ab)$. Then, by the definition of k-minimal, N^*_k is of degree $k-1$. But since N^*_k contains N'_{k-1}, the latter is a $(k-1)$-descendant of the former. Then the argument given above shows that $N^*_k - N'_{k-1} = N_k - (ab)$ is not connected.

The argument proceeds inductively without difficulty, since the last argument is independent of k. When we get to the case N'_2, $N'_1 = N_1$ is a descendant of N'_2 and thus is minimal rather than simply 1-minimal.

Condition (iv) is satisfied by our choice of the N_i.

Finally, the connected subnetworks of N_j, $2 \leq j \leq k$ are minimal. For suppose S is a connected subnetwork of N_j such that $S - (ab)$ is a connected subnetwork of S. Let $N^*_j = N'_j - (ab)$. Since N'_j is j-minimal, N^*_j is of degree $j-1$, so there exists a set of $j-1$ links whose removal from N^*_j will result in a complete disconnected subnetwork. At least one of these links is in S, since there exists, in S, a chain $(ab, q) \neq (ab)$. Thus there are at most $j - 2 \geq 0$ of these links not in N_j, so that the removal of at most $j-2$ links from the descendant N'_{j-1} results in a complete disconnected subnetwork. This is in contradiction to Lemma 2. 3 which shows that N'_{j-1} is $(j-1)$-minimal. Thus S is minimal. If $k \geq 1$, N_1 is minimal, and therefore the connected subnetworks are minimal. It follows immediately that these N_i are 1-uniform.

In the sense of this theorem, the study of an arbitrary network has been reduced to the study of a collection of 1-minimal networks. These 1-minimal networks are either connected, and so minimal, or disconnected. But a disconnected network consists of isolated nodes, isolated chains, and connected pieces. For k of the subnetworks, part (v) shows that the connected pieces are minimal. If the theorem is applied repeatedly to the connected pieces of N_{k+1}, it may, in the same sense, be reduced to isolated nodes, isolated chains, and minimal subnetworks. Thus we may say that, in a sense, the study of any network may be reduced to the study of minimal networks. This exaggerates the present state of the art, since we do not know whether this decomposition is sufficiently strong to allow general conclusions about networks, or even k-minimal networks, from a knowledge of minimal networks. In fact, an important unsolved problem is the relationship between two distinct decompositions of this type for a given network. That two distinct decompositions may exist is shown by:

$$\begin{pmatrix} 01000 \\ 00101 \\ 01010 \\ 01100 \\ 10010 \end{pmatrix} = \begin{pmatrix} 00000 \\ 00000 \\ 01000 \\ 00100 \\ 00010 \end{pmatrix} + \begin{pmatrix} 01000 \\ 00101 \\ 00010 \\ 01000 \\ 10000 \end{pmatrix} = \begin{pmatrix} 00000 \\ 00100 \\ 00010 \\ 01000 \\ 00000 \end{pmatrix} + \begin{pmatrix} 01000 \\ 00001 \\ 01000 \\ 00100 \\ 10010 \end{pmatrix}.$$

On the basis of the preceding remarks we are led to devote the rest of this paper to beginning a study of minimal networks. Section 3 includes a decomposition of any minimal network and the deduction of several properties of minimal networks. These properties are used in section 4 to draw some conclusions about arbitrary connected networks. In section 5 we discuss the relationships between several of our concepts and that of a tree in graph theory. Finally, in the last section, we present an interesting inequality, and, from this, define a subclass of minimal networks, the members of which are shown to have a particularly simple form.

3. A decomposition theorem for minimal networks. This section presents a decomposition theorem for any minimal network, which may be used to show that there exists a close connection between the concept of a minimal network and the concept of a tree in graph theory. We note first that a network which is a tree is minimal. However, the class of minimal networks is much wider than that, for we know that every connected network has a descendant, which is minimal, and we have

LEMMA 3.1. *If N is a connected network and T a descendant of N, then T is a tree only if $T = N$.*

Proof. If T is a tree and $T \neq N$, then T must have been formed from N by the removal of at least one link. Reintroduce one of these, say (ab), into T. This must introduce an oriented circuit on at least three nodes, since T is connected; let it be $(ab)(bc_1)\cdots(c_qa)$. Since this circuit is on three or more nodes, and the links of T are members of arcs, it follows that $(ab, q) = (ac_q)(c_qc_{q-1})\cdots(c_1b)$, $q \geq 2$, exists. Because of the existence of the circuit, this chain may be removed to result in the complete connected subnetwork N'. Now since $q \geq 2$, $p(N - N') > p(N - T)$, so T is not a descendant of N, which is contrary to assumption.

To carry further the work of this section we need two more definitions. First, a network is a *compound circuit of order* 1 if it is simply a non-reflexive oriented circuit on its nodes; assuming a compound circuit of order $s - 1$ defined, a *compound circuit of order* s is formed from one of order $s - 1$ by replacing some node c of that network by a non-reflexive circuit C, each link of the form (ac) by one and only one link of the form (ac') where $c' \varepsilon C$, and each link of the form (ca) by one and only one link of the form $(c''a)$, $c'' \varepsilon C$. We shall refer to this as an *inductive composition* of a compound circuit. Obviously any compound circuit is connected; furthermore, we have

12

LEMMA 3. 2. *If $N(m, p)$ is a compound circuit of order s, then $s = p - m + 1$; i. e., if N is formed by orienting the graph G, s is the first Betti number of G.*[4]

Proof. Let N be formed by the inductive composition of the circuits C_i, $i = 1, 2, \cdots, s$, in their natural order, and suppose each C_i has m_i nodes and hence the same number of links. It follows by a simple induction that

$$m = m_1 - 1 + m_2 - 1 + \cdots + m_s = \Sigma m_i - (s - 1)$$

and $p = \Sigma p_i = \Sigma m_i$, so that $s = p - m + 1$. It is well known that $p - m + 1$ is the first Betti number of any connected graph having p branches and m nodes.

A connected network N will be said to be *reducible* into subnetworks N_1 and N_2 if:

i. N_1 and N_2 are disjoint,

ii. N_1 and N_2 are each either connected or consist of a single node,

iii. there exists a network N', formed of N_1 and N_2 joined by exactly one link from N_1 to N_2 and exactly one from N_2 to N_1, such that $N' = N$.

If a connected network is not reducible it is called *irreducible.*

THEOREM 3. 3. *A connected network which is a graph is reducible if and only if it is of degree* 1.

Proof. It is clear that any reducible network is of degree 1, since we may disconnect it by removing either of the links joining the disjoint subnetworks.

Let N be a graph of degree 1 and (ab) a link such that $N - (ab) = N'$ is disconnected. Evidently, in N' there is no chain from a to b. Define M_b to consist of b and any nodes b' such that there is a chain from b' to b in N'. Let $M_a = M - M_b$. Clearly, $a \varepsilon M_a$. For any $a' \varepsilon M_a$, $a' \neq a$, there exists in N' a chain from a to a'. If not, then, since N is connected, any chain in N from a to a' must contain the link (ab), and at least one such chain exists. Since the node a can appear only one, this chain must be of the form $(aa', q) = (ab)(ba', q - 1)$, and $(ba', q - 1)$ does not contain a. But N is a graph, so $(ba', q - 1)$ implies the existence of a chain $(a'b, q - 1)$ which does not contain (ab). This, then, is a chain in N', and so $a' \varepsilon M_b$, which is

[4] König, *ibid.*, first Betti number = Zusammenhangszahl, p. 53; Whitney, *op. cit.*, first Betti number = nullity, p. 340.

contrary to choice. Thus we know a chain exists in N' from a to a'. Since N is a graph, this implies that the largest subnetworks of N defined on M_a and M_b are each either connected or consist of a single node. Now, if $a' \varepsilon M_a$ and $b' \varepsilon M_b$, where $a' \neq a$ or $b' \neq b$, then there exists no link of the form $(a'b')$, for otherwise there would be a chain from a to b in N'. Since N is a graph, it follows that there are no links of the form $(b'a')$. Thus N is reducible.

The following is our principal result:

THEOREM 3.4 (*second decomposition theorem*). *To any minimal network N, which is not a tree, there exist integers $t \geq 1$ and $y \geq 0$, such that N consists of t disjoint irreducible compound circuits C_i, $i = 1, 2, \cdots, t$, and y nodes C_{i+t}, $i = 1, 2, \cdots, y$, not included in the C_i, $1 \leq i \leq t$, subject to the conditions:*

i. *there exists at most one link from any C_i to any C_j, $i \neq j$, $1 \leq i$, $j \leq t + y$;*

ii. *no arc is contained in any of the C_i, $1 \leq i \leq t$;*

iii. *the network formed by treating the C_i, $1 \leq i \leq t$, as nodes, all other nodes and links remaining unchanged, is a tree, or, if $t = 1$ and $y = 0$, a single node.*

Proof. This proof will be carried out in two stages. First we shall show that if N is a minimal network in which there exists an arc ab, then N is reducible into two subnetworks joined only by ab. Since N is minimal, it is non-reflexive, so that $a \neq b$. Define the set of nodes M_a to consist of a and any other nodes, a', of N such that there exists a chain from a to a' which does not include the link (ab). Let $M_b = M - M_a$. $b \varepsilon M_b$, for if not, then $b \varepsilon M_a$, and so there exists (ab, q) not including (ab). Then $N - (ab)$ is a connected subnetwork of N; this violates the condition that N is minimal.

We shall now show some properties N must satisfy which will lead ultimately to a proof of the statement:

1. If $a' \varepsilon M_a$, $a' \neq a$, there exists a chain from a' to a not including the link (ba). Clearly some chain exists from a' to a, since N is connected. If all such chains include (ba), then, since the node a may only appear once, each of them may be written in the form $(a'a, q) = (a'b, q-1)(ba)$. Moreover, $(a'b, q-1)$ does not include (ab) since $a' \neq a$. Now, by the definition of M_a, there exists a chain (aa', u) which does not include (ab), so that $(aa', u)(a'b, q-1)$ does not include (ab). This is contrary to the assumption that N is minimal.

2. Let $b' \neq b$. $b' \varepsilon M_b$ if and only if there exists a chain from b to b' which does not include the link (ba). Suppose first that $b' \varepsilon M_b$. Since N is connected, there exists at least one chain from b to b'. Suppose each (bb', q) contains (ba). Then, since each node may appear only once, $(bb', q) = (ba)(ab', q-1)$. If $(ab', q-1)$ does not contain (ab), then, by definition, $b' \varepsilon M_a$, which is impossible. But $(ab', q-1)$ cannot contain (ab), for if it did, then (bb', q) would not be a chain. Hence (bb', q) does not contain (ba).

Conversely, suppose there exists a chain from b to b' not including (ba). If $b' \varepsilon M_a$, there exists, by 1, a chain from b' to a not including (ba); these two combine into a chain from b to a not including (ba), which is contrary to N being minimal.

3. If $b' \varepsilon M_b$, $b' \neq b$, there exists $(b'b, q)$ not including (ab). We may parallel the proof of property 1 by replacing the words "definition of M_a" by "property 2."

4. If $a' \varepsilon M_a$, $b' \varepsilon M_b$, and either $a \neq a'$ or $b \neq b'$, then no link of the form $(a'b')$ exists. Suppose such a link does indeed exist. Then, by the definition of M_a, there exists a chain (aa', q) which does not include (ab), and, by property 3, a chain $(b'b, r)$ which does not include (ab). Thus the chain $(aa', q)(a'b')(b'b, r)$ does not include (ab), since $(a'b') \neq (ab)$, which is impossible.

5. Under the same conditions as in 4, there is no link of the form $(b'a')$. The argument is exactly parallel to that of 4, using properties 1 and 2.

It thus follows that the maximal subnetworks of N on the sets M_a and M_b are each either connected and minimal, or consist of a single node. From 4 and 5, one concludes that the subnetworks are joined only by the arc ab. This exhausts N, and the result is proved.

Since an arbitrary network has a finite number of arcs, it follows from a finite number of applications of the above result that a minimal network which is not a tree consists of a set of $t' \geq 1$ disjoint arc-free minimal subnetworks C_i, $i = 1, 2, \cdots, t'$, and $y' \geq 0$ nodes $C_{t'+i}$, $i = 1, 2, \cdots, y'$, not included in the C_i, $1 \leq i \leq t'$, such that:

i. any link not in a C_i, $1 \leq i \leq t'$, is a member of an arc;

ii. there exists at most one arc between any C_i and C_j, $i \neq j$, $1 \leq i$, $j \leq t' + y'$;

iii. the network formed by treating the C_i, $1 \leq i \leq t'$, as nodes, is a tree;

iv. the decomposition is unique.

By virtue of this decomposition, the problem is reduced to examining the case of an arc-free minimal network. We show: An arc-free minimal network consists of $t'' \geq 1$ disjoint irreducible compound circuits C_i, $i = 1, 2, \cdots, t''$, and $y'' \geq 0$ nodes $C_{i+t''}$, $i = 1, 2, \cdots, y''$, not included in the C_i, $1 \leq i \leq t''$, such that:

i. there exists at most one link from any C_i to any C_j, $i \neq j$, $1 \leq i$, $j \leq t'' + y''$;

ii. the network formed by treating the C_i, $1 \leq i \leq t''$, as nodes is a tree, or, if $t'' = 1$ and $y'' = 0$, a single node.

The first arc-free minimal network occurs for $m = 3$, and this obviously satisfies the conditions since it is a circuit on three nodes. Suppose now that the statement, except for the condition that the compound circuits are irreducible, has been proved for all networks through $m - 1$ nodes, and let $N(m, p)$ be an arc-free minimal network. In N there exists a circuit consisting of at least three links, since N is connected, arc-free, and non-reflexive; let C be one such circuit on the nodes c_i, $i = 1, 2, \cdots, q \geq 3$. The maximum subnetwork of N on these nodes is only C, for if there exists any other link $(c_i c_k)$, $k \neq i + 1$, then this link can be removed without disconnecting N, since the chain $(c_i c_{i+1})(c_{i+1} c_{i+2}) \cdots (c_{k-1} c_k)$ exists. This is impossible since N is minimal. Now, if C exhausts all the nodes of N we are done. If not, let M' be the set of nodes remaining. If $a \in M'$, and (ac_k), $1 \leq k \leq q$, exists, then no link of the form (ac_i), $1 \leq i \leq q$, $i \neq k$, exists. For if so, then the chain from c_k to c_i, which is a part of C, and so does not include (ac_i), shows that (ac_i) may be removed without disconnecting N. This is impossible. Similarly, if $(c_k a)$, $1 \leq k \leq q$, exists, then $(c_i a)$, $1 \leq i \leq q$, $i \neq k$, does not exist.

Now consider the network N' formed by letting the nodes of C coalesce into a single node which we shall call c. Evidently, since N is minimal, so is N', and N' has at least two nodes fewer than N, since $q \geq 3$. Several possibilities exist for N'. First, it may be a graph, and hence a tree, in which case the statement is proved. Second, it is not a tree, but there exists at least one arc. By the first part of this theorem, N' may be decomposed into several arc-free minimal subnetworks connected in such a fashion that if they are treated as nodes, the resulting network is a tree. By the induction hypothesis, these arc-free minimal subnetworks satisfy the conditions of the statement we are proving. But replacing the node c by C, reconstructing N, only increases the order of one of these compound circuits, or introduces a new compound circuit, so the result is true for N. Third, N' is arc-free, in

which case the induction hypothesis may be applied directly, and the introduction of C for c only increases the order of the compound circuit.

Thus, we may decompose N into several compound circuits and nodes not in these compound circuits and connecting links satisfying the conditions i and ii of the second intermediate statement. Carry this decomposition as far as possible; the process will terminate in a finite number of steps, since N is finite. We will show that the resulting compound circuits are irreducible. For suppose that C_k is reducible into the disjoint subnetworks A and B connected by the links (ab), $(b'a')$, $a, a' \varepsilon A$, $b, b' \varepsilon B$. By condition ii it follows that any C_i, $i \neq k$, $1 \leq i \leq t' + y'$ is linked "symmetrically" to C_k if at all. In fact, it is either linked symmetrically to A or to B; for if not, then there exists a link from A to C_i and a link from C_i to B, in which case (ab) may be removed, or, in the other case, $(b'a')$ may be removed without disconnecting N. This is impossible. A and B are either compound circuits or, by the result proved for arc-free minimal networks, may be reduced to several compound circuits and nodes not in them such that i and ii hold. By an argument similar to the one just made, the conditions i and ii hold for N with this finer decomposition. This is contrary to choice, so C_k must be irreducible.

The proof of the theorem follows almost immediately from the two intermediate results, if we note that the last argument may be applied to show condition iii.

This decomposition of a minimal network is not unique, for

$$\begin{pmatrix} 010000 \\ 000001 \\ 010100 \\ 000010 \\ 001000 \\ 100010 \end{pmatrix}$$

may be decomposed into either a tree consisting of one arc, or one of two arcs.

The next result gives a little more information about the components into which we have decomposed a minimal network, the irreducible compound circuits. This result is unsatisfactory in the sense that it does not give a complete characterization of these networks. For this proof and succeeding results we need the following definition. A node is *simple* if it is the initial node of exactly one link and the end node of exactly one link.

THEOREM 3.5. *Let N be a minimal network. N is irreducible if and*

only if it is a compound circuit such that in any inductive composition of N, none of the circuits introduced are arcs.

Proof. Suppose that at some stage of the composition of N, an arc ab is introduced into a compound circuit C to form a compound circuit C'. If ab does not have a simple node, then in C' there exists either a chain from a to b not containing (ab), or one from b to a not containing (ba), since such a chain exists in C. The introduction of further circuits can only lengthen this chain, so $N - (ab)$ is a complete connected subnetwork, which is impossible. Hence a node of ab is simple. The introduction of further circuits merely adds to C to form a larger compound circuit, and hence a connected subnetwork or a single node of N, or it may replace the simple node of ab by a compound circuit. Between these two connected subnetworks, or single nodes, are only the links arising from ab, now no longer an arc in general. Thus N is reducible, which is contrary to assumption, proving that no arc can be introduced.

Conversely, if we suppose N is reducible, then Theorem 3.4 implies N may be decomposed into one or more irreducible compound circuits and nodes not included in these compound circuits. The circuits of any compound circuit C may be coalesced into nodes in the inverse order of an inductive composition of C. This clearly leads to the tree of Theorem 3.4. But any tree is a compound circuit formed only of arcs. Thus we have an inductive composition of N involving arcs, the arcs of the tree. As this is contrary to assumption, N must be irreducible.

The principal theorem will be utilized sometimes through two properties of minimal networks derivable from it. They are presented as

THEOREM 3.6. *A minimal network is a compound circuit which contains at least two simple nodes.*

Proof. The last part of the above proof suffices to show that a minimal network is a compound circuit.

To show that a minimal network has two simple nodes, we shall perform an induction on the order s of the compound circuit. It is certainly true for $s = 1$, since the compound circuit is then a non-reflexive circuit. Assume the result true up through circuits of order $s - 1$. Suppose N is a minimal compound circuit of order s, and let C be the last circuit introduced in some inductive composition of N. Let C coalesce into a single node c, and call the resulting network N'. N' is readily seen to be minimal and of order $s - 1$,

so by the induction hypothesis it contains at least two simple nodes. If two of these are different from c, then we are done. If not, c is simple. Consider N; if C is not an arc then it must introduce a simple node, for C has at least three nodes, and there exists only one link to C from the rest of the nodes, and only one from C, since c is simple. If, on the other hand, C is an arc, then the first argument in the proof of Theorem 3.5 shows that one of its nodes is simple, and the result follows.

That not every compound circuit is minimal or has a simple node is shown by:

$$\begin{bmatrix} 0101 \\ 0010 \\ 0101 \\ 1000 \end{bmatrix}.$$

4. Applications to connected networks. Two applications to connected networks are given of the results on minimal networks; the first examines limits on the number of links a connected network may have, and the second discusses the maximum number of "independent" circuits a connected network may have.

THEOREM 4.1. *Let $N(m,p)$ be a connected network, not a tree. Let N have a descendant N' which is decomposable in the terms of Theorem 3.4 into t irreducible minimal subnetworks and y nodes not in these subnetworks. Then*

$$p \leq (3m + t + y - 4)/2 + p(N - N') < 2(m - 1) + p(N - N').$$

If N is a tree, $p = 2(m - 1)$.

Proof. If N is a tree, the result is well known from graph theory.[5]

Suppose N is not a tree. Then it is sufficient to show the result for the class of minimal networks which are not trees, since, in the general case, the network N has $p(N - N')$ more links than any descendant N'. By Lemma 2.3 a descendant is minimal, and, by Lemma 3.1, it is not a tree. So we consider N minimal. Decompose N as in Theorem 3.4, and let the irreducible compound circuits C_i have m_i nodes, p_i links, and order s_i. Let there be p' links not in any irreducible compound circuit. Then, by the result on trees, $p' = 2(t + y - 1)$. For each of the irreducible minimal subnetworks, Theorem

[5] Whitney, *ibid.*, pp. 340-341.

3.5 implies that each of the s_i circuits used in forming C_i has at least three nodes, so that $m_i \geq 3 + 2 + 2 + \cdots + 2 = 2s_i + 1$. By Lemma 3.2,

$$p_i = s_i + m_i - 1 \leq 3(m_i - 1)/2.$$

Thus,

$$p = \Sigma p_i + p' \leq \Sigma 3(m_i - 1)/2 + 2(t + y - 1)$$
$$= (3/2)(\Sigma m_i + y) + (t + y - 4)/2 = (3m + t + y - 4)/2.$$

This may be simplified a little by noting that each of the irreducible minimal subnetworks must have at least three nodes, so $m \geq 3t + y$; hence,

$$p \leq (3m + t + y - 4)/2 = (4m - 4 - 2t + 3t + y - m)/2$$
$$\leq 2(m - 1) - t < 2(m - 1).$$

This concludes the proof.

It is clear that in a given network we may define the *addition of chains* (*mod* 2). Thus we may also define *linear independence* (*mod* 2). We shall be concerned with sets of linearly independent (mod 2) circuits such that no other linearly independent set contains a greater number of circuits. These sets will be called *maximal*. The result proved in the next theorem is, in statement, formally the same as a result of graph theory:[6]

THEOREM 4.2. *In any connected network $N(m, p)$ there exists a maximal set of $p - m + 1$ linearly independent* (*mod* 2) *circuits.*

Proof. First, it is sufficient to show this for minimal networks. For, if N is not minimal, then it has a descendant N' which is. N may be considered to be formed from N' by the addition of links one at a time. Each such link adds at least one new circuit which is independent (mod 2) of the circuits of the network to which it was added, since in a connected network every link is contained in at least one circuit. Thus, if there exists a set U' of $p - K - m + 1$, $K = p(N - N')$, linearly independent (mod 2) circuits in N', there will exists a set U of at least $p - m + 1$ linearly independent circuit in N.

Furthermore, if U' is maximal in the descendant, U will be in N also. If not, then there is a first subnetwork, N^*, for which any set of $p^* - m + 1$ linearly independent (mod 2) circuits is maximal, and to which the addition of a link (ab) produces a linearly independent set U'' having more than

[6] Lefschetz, Solomon, *Introduction to Topology*, Princeton, Princeton University Press (1949), p. 71.

$p^* - m$ circuits. It is clear that this set U'' must contain at least two circuits which include the link (ab), for otherwise the subset of U'' in N^* would contain more than $p^* - m + 1$ linearly independent circuits. Let two of the circuits be denoted by $(ab)(ba, q)$ and $(ab)(ba, q')$. Since N^* is connected and does not contain (ab), there exists a chain from a to b not including (ab); select a shortest: (ab, q''), $q'' > 1$. In general, (ab, q'') will coincide with (ba, q) over a certain number of links, i. e., over a set of several chains of the form (cd, t), each a part of (ba, q). The argument does not change in principal, and a great saving in notation is gained, if we assume that at most one such chain occurs. Similarly, (ab, q'') will be assumed to coincide with (ba, q') over the chain $(c'd', t')$. Furthermore, we shall assume that (cd, t) and $(c'd', t')$ have no links in common; if they do, a slight modification of the following argument will suffice. So we may write:

$$(ab, q'') = (ac, u)(cd, t)(dc', z)(c'd', t')(d'b, v),$$

the order of (cd, t) and $(c'd', t')$ being immaterial.

$$(ba, q) = (bc, x)(cd, t)(da, y), \qquad (ba, q') = (bc', x')(c'd', t')(d'a, y').$$

Observe that the following formal products are in fact circuits of N^*:

A: $(ac, u)(cd, t)(da, y)$ B: $(d'b, v)(bc', x')(c'd', t')$

C: $(ac, u)(cd, t)(dc', z)(c'd', t')(d'a, y')$

D: $(bc, x)(cd, t)(dc', z)(c'd', t')(d'b, v)$.

Since these are circuits of N^*, they are expressible $(\text{mod } 2)$ in terms of the circuits in U''. But observe that $(ab)(ba, q) + A + B + C + D = (ab)(ba, q')$ $(\text{mod } 2)$ so that $(ab)(ba, q)$ and $(ab)(ba, q')$ are not linearly independent. Thus only one of them can be in U'', and so we have shown that if the theorem is true for minimal networks it is true in general.

The minimal case will be proved by induction on m. For $m = 2$ it is trivially true. Suppose it is true for all minimal networks having $m - 1$ or fewer nodes, and let $N(m, p)$ be minimal. By Theorem 3. 6, N has a simple node a which is the initial node of only one link, (ab), and the end node of only one, (ca). We may distinguish three cases:

i. $b = c$. Remove a and the arc ab, leaving the subnetwork $N'(m - 1, p - 2)$. N' is obviously minimal, and so it has a maximal set of $p - m$ linearly independent $(\text{mod } 2)$ circuits. The arc ab adds exactly one circuit to this set.

ii. $b \neq c$, and there does not exist $(cb, q) \neq (ca)(ab)$. Remove a and the adjoining links and introduce the link (cb) to form $N'(m-1, p-1)$, which is minimal. Thus, by the induction hypothesis, has a maximal set of $p-m+1$ linearly independent circuits. But forming N' from N cannot essentially change any set of linearly independent circuits, since the chain $(ca)(ab)$ is, in this case, formally the same as (cb).

iii. $b \neq c$, and there does exist $(cb, q) \neq (ca)(ab)$. Again remove a and the adjoining links to form $N'(m-1, p-2)$, which is minimal. By the induction hypothesis, N' has a maximal set of $p-m$ linearly independent circuits. Replacing a and the two links (ab) and (ca) to form N adds one or more new circuits, depending on the number of chains from b to c. This situation is not essentially different from the one discussed in the first part of this proof, except that we are adding a 2-chain and a new node, rather than a single link. Since this node is simple, the argument is formally the same, and it shows that there is a maximal set of $p-m+1$ linearly independent (mod 2) circuits in N. This, then, concludes the proof.

We note the trivial corollary: A connected network $N(m, p)$ has exactly $p-m+1$ circuits if and only if the set of all circuits of N is linearly independent (mod 2).

5. Generalizations of a tree. We shall show in this section that several of our definitions, when applied to networks which are graphs, are identical with the concept of a tree. This can be shown directly and easily in each case; however we shall first prove two results which are true in general, and then we shall use them to prove Theorem 5. 3. Thus, that result is not as deep as it first appears to be.

We shall call a connected network $N(m, p)$ having exactly $p-m+1$ circuits *circuit minimal*. This definition makes sense because of Theorem 4. 2. By the corollary to that theorem, N is circuit minimal if and only if N is connected and the set of all circuits is linearly independent (mod 2).

LEMMA 5. 1. *A circuit minimal network is uniform.*

Proof. Let $N(m, p)$ be circuit minimal. Let (ab) be any link, and $N-(ab)=N'$. If N' is not connected, then N has degree 1. If N' is connected it is circuit minimal, for at least one circuit of N was destroyed by the removal of (ab), and according to Theorem 4. 2, no more than one.

Since N' is connected, there exists at least one chain from b to a, but only one, for if there were more then the addition of the single link (ab) would introduce more than one circuit, and N would have more than $p-m+1$ circuits. In N, interrupt the chain from b to a by removing a single link from it, thus disconnecting N. This proves N is of degree 1.

To show N is uniform it will thus suffice to show that every connected subnetwork is circuit minimal. If S is a connected network which is not circuit minimal, then there exists a circuit in S which is linearly dependent (mod 2) on the other circuits of S. This remains true in N, so N is not circuit minimal, a contradiction.

LEMMA 5. 2. *A compound circuit is uniform.*

Proof. This may be demonstrated by an induction on the order of compound circuits. It is trivially true for compound circuits of order 1. Let N be a compound circuit of order $s > 1$, let C be the last circuit introduced in some inductive composition of N, and let S be any subnetwork of N. Coalesce C into a single node c, and under this operation let S become S'. If S' is a single node, then $S = C$, and the degree of S is 1. Otherwise, S' is connected, and therefore, by the induction hypothesis, it is of degree 1. Select $(ab) \varepsilon S, \not\varepsilon C$, such that $S' - (ab)$ is not connected. This is possible since S′ is of degree 1. Now the introduction in S' of that part of C in S, subject to the conditions of N, can only replace the node c by a chain or a circuit, but cannot introduce a link or a chain from a to b; thus S is of degree 1. So N is uniform.

The next result is the justification for the title of this section.

THEOREM 5. 3. *For a connected network N which is a graph, the following are equivalent*: i. *N is a tree,* ii. *N is minimal,* iii. *N is a compound circuit,* iv. *N is uniform,* v. *N is circuit minimal.*

Proof. i. implies ii trivially. ii. implies iii by Theorem 3. 6. iii. implies iv by Lemma 5. 2. iv. implies i. For if N is not a tree, then there exists a circuit in the sense of graph theory. But this is clearly a connected subnetwork of degree 2, so that N is not uniform. v. implies iv by Lemma 5. 1. i. implies v. If $N(m, p)$ is a tree, it follows, from theorem 3. 5, that $p-(m-1) = m-1$. Furthermore, the only circuits in a tree are 2-circuits (arcs) of which there are exactly $m-1$, so the number of circuits is $p-m+1$.

The several results of this section and Theorem 3. 6 suggest the following

class of unsolved problems: Conditions on a uniform network that it be a compound circuit. Conditions on a uniform network that it be circuit minimal. Conditions on a compound circuit that it be minimal. These four concepts are indeed all distinct. The network.

$$\begin{pmatrix} 010000 \\ 001001 \\ 000100 \\ 100010 \\ 010000 \\ 000100 \end{pmatrix}$$

is minimal, and hence a compound circuit, but not circuit minimal. The network

$$\begin{pmatrix} 0101 \\ 1000 \\ 1101 \\ 1010 \end{pmatrix}$$

is both uniform and circuit minimal, but not a compound circuit. The network

$$\begin{pmatrix} 0111 \\ 1010 \\ 1001 \\ 1000 \end{pmatrix}$$

is uniform, but neither a compound circuit nor circuit minimal. The example following Theorem 3. 6 shows that not every compound circuit is minimal.

6. Rank minimal networks. In section 1 we noted the representation of networks by relation matrices with entries from the two-element Boolean algebra. Equally well, we may interpret this as a representation by real matrices with the numbers 0 and 1 as entries. Thus, since it is well known that matrix rank is a similarity invariant, to each network there is a uniquely defined number r, $1 \leq r \leq m$, called the *rank of the network*, which is the rank of any of the corresponding real matrix representations.

THEOREM 6. 1. *If $N(m, p)$ is a connected network*[7] *having rank r, then $p + r \geq 2m$.*

---

[7] Luce, R. D., "Connectivity and generalized cliques in sociometric group structures," *Psychometrika*, vol. 15 (1950), pp. 169-190. In this paper the diameter, n, of a connected network was defined as $n = \max_{a,b \in M} \min_q (ab, q)$, and it was conjectured that $p + n \geq 2m$. This is now known to be false; however, Theorem 6. 1 is a correct result which is closely related to the conjecture, for it may also be shown that $r \geq n$.

Proof. Suppose $p + r < 2m$. Select any set R of r linearly independent rows in a particular matrix representation. Since N is connected, there exists a non-zero entry in each column j; but since each row can be written as a linear combination of rows from R, it follows that for each column j there exists an $i \varepsilon R$, such that the ij entry is 1. Thus, in the rows of R there are at least m 1's. By our assumption, there remain $p' \leq p - m < m - r$ links (entries that are 1). Each of the $m - r$ rows not in R must contain a non-zero entry, since N is connected, and therefore $p' \geq m - r$, a contradiction.

We shall call a network $N(m, p)$ *rank minimal* if it is connected, and $p + r = 2m$.

THEOREM 6.2. *If a connected network is rank minimal, it is minimal.*

Proof. As in the proof of Theorem 6.1, we consider a matrix representation N of the rank minimal network N, and let R be a set of r linearly independent rows. Each column has a non-zero entry in some row of R, since N is connected. The set R' of the $m - r$ remaining rows must have a non-zero entry in each row for the same reason. However, since $p = 2m - r$, it is necessary that R have exactly one non-zero entry in each column, and R' exactly one in each row.

Let (ab) be any link of the network. We shall show that its removal results in a disconnected network, which will prove the theorem.

If $a \varepsilon R'$, then the removal of (ab) results in a network N' having no link for which a is the initial node, since the rows of R' have exactly one non-zero entry.

If $a \varepsilon R$, then either the row a has only one non-zero entry, and we use the above argument, or it has another non-zero entry, say in column c, $c \neq b$. We show that in the latter case column b has only the one non-zero entry, N_{ab}. For suppose another link (db), $d \neq a$, exists. Then $d \varepsilon R'$, for we showed above, essentially, that the rows of R have exactly one non-zero entry in each column, and we have assumed (ab) to exist and $a \varepsilon R$. Since the rows of R' have exactly one non-zero entry, it follows that N_{db} is the one for row d. But since the rows of R are a set of linearly independent ones for this matrix, row d must be a linear combination of rows of R. The row a must be in this combination, as it is the only one of R having an entry in the b column. However, we assumed that row a has a non-zero entry in column c. This must be subtracted, since row d cannot have an entry $N_{dc} = 1$. But this is impossible using only rows of R, since no other row of R has an entry in the

c column. This contradiction implies that column b has only $N_{ab} = 1$, and so $N - (ab)$ is a complete disconnected subnetwork of N. Thus N is minimal.

The converse statement is not true, as will be obvious from a comparison of Theorem 6. 4 and Theorem 3. 4.

The next lemma will be used in conjunction with Theorem 3. 4 to decompose any rank minimal network.

LEMMA 6. 3. *Let N be rank minimal. If N is reducible into the subnetworks N_1 and N_2, then either N_1 or N_2 is a single node.*

Proof. Let the N_i, $i = 1, 2$, have m_i nodes, p_i links, and rank r_i; and let m, p, and r denote the corresponding quantities in N. If neither of the N_i is a node, they are both connected subnetworks, so by Theorem 6. 1, $p_i \geq 2m_i - r_i$. It is evident from the definition of a reducible network that $p = p_1 + p_2 + 2$, and $m = m_1 + m_2$. Furthermore, if we let the matrix minor repersentation of N_i be denoted by the same symbols, we then have, for an appropirate labeling of the nodes, the following type of matrix representation for N:

$$\begin{bmatrix} & \begin{matrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{matrix} \\ \smash{\raise 4em{N_1}} & \\ \begin{matrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{matrix} & N_2 \end{bmatrix},$$

whence one sees that $r \geq r_1 + r_2 - 1$. Thus, $p = p_1 + p_2 + 2 \geq 2m_1 - r_1 + 2m_2 - r_2 + 2 = 2(m_1 + m_2) - (r_1 + r_2 - 1) + 1 > 2m - r$, which is contrary to the assumption that N is rank minimal.

A tree such that the arcs all have one end node in common is called a star. It is not difficult to show that a star is rank minimal.

THEOREM 6.4. *Let N be a rank minimal network which is not a star. If N has any arcs, there exists one arc-free rank minimal subnetwork C of N, such that the network formed from N by treating C as a node, all other links and arcs remaining unchanged, is a star. An arc-free rank minimal network C consists of exactly one irreducible rank minimal subnetwork C', not a single node, and, possibly, some simple nodes such that the network formed from C by treating C' as a node, all other links and nodes remaining unchanged, is a star. Furthermore, if C' is the irreducible rank minimal subnetwork of the arc-free rank minimal subnetwork C of N, then the network formed from N by treating C' as a node, all other links and nodes remaining unchanged, is a star.*

Proof. In the first case, apply Lemma 6.3 to the first statement presented in the proof of Theorem 3.4 to show that each arc which exists must have a simple node. The arc-free subnetwork C is rank minimal, for the removal of any arc from N results in a network N' for which $p' = p - 2$, $m' = m - 1$, and $r' \leq r$, implying $p' \leq 2m' - r'$. Thus, by Theorem 6.1, N' is rank minimal. To C, first apply Theorem 3.4 and then Lemma 6.3 to show any nodes, not in an irreducible subnetwork, must be simple, and there is only one irreducible subnetwork C'. The same argument as applied above suffices to show that C' is rank minimal. The final statement is proved in the same manner.

In conclusion one may mention two more unsolved problems: Conditions on a minimal network that it be rank minimal, and a characterization of an irreducible rank minimal network.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY.

ON THE NON-VANISHING OF CERTAIN DIRICHLET SERIES.*

By Aurel Wintner.

Let $f(n)$, where $n = 1, 2, \cdots$, be a completely multiplicative function, that is, let $f(n_1 n_2) = f(n_1) f(n_2)$ but $f(1) \neq 0$. Such a function is uniquely determined by an arbitrary assignment of the values $f(p)$, and is a bounded function if and only if $|f(p)| \leqq 1$ holds for every prime. The following theorem will be proved:

If $f(n)$ is completely multiplicative and bounded, and if the function $F(s)$, defined for $\sigma > 1$ by

$$F(s) = \sum_{n=1}^{\infty} f(n)/n^s, \tag{1}$$

has no singular point on $\sigma = 1$, then it has at most one zero on $\sigma = 1$.

This assertion is meant to imply that the zero, if any, cannot be a multiple zero. That it can occur at all, is shown by Liouville's example, $f(p) = -1$, where $s = 1$ is a zero of $F(s) = \zeta(2s)/\zeta(s)$. Since this $F(s)$ is replaced by $F(s - ia)$ if every $f(p) = -1$ is multiplied by p^{ia}, the zero can occur at any point, $s = 1 + ia$, of the line $\sigma = 1$.

For reasons of symmetry, the zero, if any, must be at $s = 1$ if $f(n)$ is real-valued. In this case, the assumption of boundedness, which is then equivalent to $-1 \leqq f(p) \leqq 1$, can be refined to $-1 \leqq f(p)$, if (1) is absolutely convergent for $\sigma > 1$. This was proved in [2] by an argument based on $\zeta(s)F(s)$. The above theorem will be proved by combining that argument with a device, introduced in this context by Ingham [1], which replaces $\zeta(s)F(s)$ by

$$G(s) = \zeta^2(s) F(s) F^*(s), \tag{2}$$

where $F^*(s)$ denotes the Dirichlet series the coefficients of which are the complex conjugates of the coefficients of (1).

First, if $\sigma > 1$, then, since $|f(n)| \leqq 1$, logarithmic differentiation of the Euler factorization of (1) gives

* Received June 15, 1951.

(3) $$-F'/F(s)=\sum_{n=1}^{\infty}\Lambda(n)f(n)/n^s, \text{ where } -\zeta'/\zeta(s)=\sum_{n=1}^{\infty}\Lambda(n)/n^s.$$

Hence, from (2),

(4) $$-G'/G(s)=\sum_{n=1}^{\infty}2\{1+\Re f(n)\}\Lambda(n)/n^s.$$

Since $F(s)$ is supposed to remain regular on $\sigma=1$, it is clear from (2) that the limit

(5) $$m=m(t)=\lim_{\epsilon\to 0}\epsilon G'/G(1+\epsilon+it)$$

exists for every real t and is the order of the zero $s=1+it$ of $G(s)$, with the understanding that this order can be negative or 0. On the other hand, since $|f(n)|\leqq 1$ and $\Lambda(n)\geqq 0$, every coefficient of (4) is non-negative. Hence it is clear from (5) that $|m(t)|\leqq|m(0)|$ holds for every t.

In particular $m(t)$ must vanish identically if it vanishes at $t=0$. This means that $G(s)$ must be regular and non-vanishing at every $s=1+it$ if it is regular and non-vanishing at $s=1$. But the latter assumption is satisfied if $F(s)$ vanishes at $s=1$ in the first order. This is clear from (2), since $\zeta(s)$ has a pole of first order at $s=1$. Consequently, if $F(s)$ has a simple zero at $s=1$, then $G(s)$ has no zero $s=1+it\neq 1$, which, in view of (2), means that $s=1$ is the only zero of $F(s)$ on the line $\sigma=1$. It follows that, in order to prove that

(6) $$F(1+it)\neq 0 \text{ for every } t\neq 0 \text{ if } F(1)=0,$$

it is sufficient to show that $F(s)$ cannot have a multiple zero at $s=1$.

Since $|f(n)\Lambda(n)|\leqq\Lambda(n)$, it is clear from (3) that

$$|F'/F(1+\epsilon)|\leqq|\zeta'/\zeta(1+\epsilon)|$$

if $\epsilon>0$. It follows therefore from

$$\lim_{\epsilon\to 0}\epsilon\zeta'/\zeta(1+\epsilon)=-1 \text{ that } \lim_{\epsilon\to 0}|\epsilon F'/F(1+\epsilon)|\leqq 1.$$

Finally, the last inequality implies that $F(s)$ cannot have a multiple zero at $s=1$, i. e., that

(7) $$F'(1)\neq 0 \text{ if } F(1)=0.$$

This proves (6). But (6) implies that

(8) $$F(1+it)\neq 0 \text{ for every } t\neq a \text{ if } F(1+ia)=0,$$

where a is any real number. In fact, (8) follows if $f(n)$ in (1) is replaced by $f(n)n^{-ia}$ and then (6) is applied to the new function (1). Similarly, (7) implies that

$$(9) \qquad F'(1+ia) \neq 0 \text{ if } F(1+ia) = 0.$$

Clearly, (8) and (9) together are equivalent to the theorem italicized above.

It is clear from the proof that, instead of assuming the regularity of $F(s)$ on $\sigma = 1$, it is sufficient to assume that $F(s)$ and $F'(s)$, where $\sigma > 1$, go over into continuous boundary values as $\sigma \to 1$. In fact, a somewhat less stringent condition would also suffice.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] A. E. Ingham, "Note on Riemann's zeta-function and Dirichlet's L-functions," *Journal of the London Mathematical Society*, vol. 5 (1930), pp. 107-112.

[2] A. Wintner, "On the non-vanishing of certain Dirichlet series," *Rendiconti del Circolo Matematico di Palermo*, New series, vol. 1 (1952).

ON THE FUNDAMENTAL GROUP OF AN ALGEBRAIC VARIETY.*

By Wei-Liang Chow.

It is well known that any 1-cycle in an algebraic surface can be deformed into a 1-cycle lying in a generic plane section of the surface.[1] The usual proof of this theorem, which can be easily generalized from a surface to any non-singular algebraic variety, is topological and consists of a simple construction of the deformation chain. In the transcendental theory there is a generalization of this theorem, at least in its homology aspect, which can be stated as follows:[2] There exist exactly $2p$ independent 1-cycles in an algebraic surface which are not homologous to 1-cycles belonging to a generic curve of an irrational pencil of genus p. In this paper we shall show that this theorem is a special case of a more general theorem about the fundamental group of an algebraic variety under a rational transformation. Our method of proof will be purely topological; the essential idea is that although a rational transformation is not in general a fibre mapping, the covering homotopy theorem is nevertheless true, in a somewhat modified form, for the mapping of a 1-simplex.

In section 1 the notion of a fibre system is introduced, and certain subsystems of an algebraic system are shown to be (or can be considered as) fibre systems. The notion of a fibre system is a generalization of that of a fibre space, and just as in the case of a fibre space we have also here as a fundamental property the validity of the covering homotopy theorem, which must now be formulated in a somewhat modified form. This notion of a fibre system is a very useful tool in the study of the topology of algebraic varieties; in this paper we shall limit ourselves strictly to the particular problem in question, but we hope to show in a later paper that the method is applicable also to other similar problems in algebraic geometry. In section 2 the results of section 1 will be used to prove two theorems; one of them (Theorem 2) is the theorem mentioned above, the other (Theorem 1) is a theorem concerning the deformation of 1-cycles into a member of an algebraic system with at least

* Received August 20, 1951.

[1] See, e. g., O. Zariski, *Algebraic Surfaces*, p. 108.

[2] See O. Zariski, *Algebraic Surfaces*, p. 144.

one base point, which can also be regarded as a (partial) generalization of a result of Severi[3] (proved by transcendental methods).

1. Let U and V be topological spaces, and let $G(y)$ be a function which assigns to each point y in V a subset $G(y)$ in U. We shall say that the system of subsets $G(y)$ defines a fibre system in U, if there exists an open covering $N = \{N\}$ of V such that for each set N there exists a continuous function $\phi_N(x, y)$, defined for all points $x \subset G(N)$, $y \subset N$ in the product space $U \times V$ with values in U, with the following properties:

$$\phi_N(x, y) \subset G(y), \qquad (x \subset G(N), y \subset N),$$

$$\phi_N(x, y) = x \qquad (x \subset G(y), y \subset N).$$

The space V is called the base space of the fibre system, and the open sets N and the corresponding functions $\phi_N(x, y)$ are called the slicing neighborhoods and the slicing functions respectively. Let $f(z)$ and $g(z)$ be continuous mappings of a topological space Z into U and V respectively, such that for each point $z \subset Z$ we have $f(z) \subset G(g(z))$, and let $g(z, t)$, $0 \leq t \leq 1$, be a homotopy of the mapping $g(z)$ in V. Then a homotopy $f(z, t)$, $0 \leq t \leq 1$, of $f(z)$ in U is said to cover the homotopy $g(z, t)$, or a covering homotopy of $g(z, t)$, if we have $f(z, t) \subset G(g(z, t))$ for all z, t. The covering homotopy theorem (in the weak form) asserts that if V is a normal Hausdorff space and if Z is compact, then there exists always a covering homotopy; furthermore, if $g(z, t)$ leaves a point $z_0 \subset Z$ fixed, we can assume that $f(z, t)$ also leaves z_0 fixed. That this covering homotopy theorem is true for any fibre system in U can be seen as follows. We observe first that in case the $G(y)$ is the inverse function $\pi^{-1}(y)$ of a continuous mapping $\pi(x)$ of U onto V, then we have a (generalized) fibre space as defined by S. T. Hu[4]; and, as has been observed by Hu, the covering homotopy theorem is true for such a fibre space. The general case can be reduced to this special case by considering the graph W of the function $G(y)$, i. e. the set of all points $w = x \times y$ in $U \times V$ satisfying the condition $x \subset G(y)$. Let $\tau(w)$ and $\pi(w)$ be the

[3] F. Severi, "Intorno al teorema d'Abel sulle superficie algebriche ed alla riduzione a forma normale degl'integrali di Picard," *Rendiconti del Circolo Matematico di Palermo*, vol. 21 (1906), p. 261, Teorema I.

[4] Sze-Tsen Hu, "On generalizing the notion of fibre spaces to include the fibre bundles," *Proceedings of the American Mathematical Society*, vol. 1 (1950), pp. 756-762. It is convenient to use this generalized notion of a fibre space, though the particular fibre systems used in the present paper can all be "derived" from fibre bundles which are also at the same time fibre spaces in the sense of Hurewicz-Steenrod.

mappings of W into U and V induced by the projections of $U \times V$ into U and V respectively. It is clear that $\pi(w)$ is a mapping of W onto V, and that we have $\tau(\pi^{-1}(y)) = G(y)$ for each $y \subset V$. The space W can then be made into a fibre space with respect to the mapping $\pi(w)$ if we define for each N the slicing function by the formula:

$$\phi_N(w, y) = \phi_N(\tau(w), y) \times y \qquad (w \subset \pi^{-1}(N), y \subset N).$$

If we set $h(z) = f(z) \times g(z)$, then $h(z)$ is a continuous mapping of Z into W such that $h(z) \subset \pi^{-1}(g(z))$. Since the covering homotopy theorem holds for the fibre mapping $\pi(w)$, there exists a homotopy $h(z, t)$ of $h(z)$ in W such that $h(z, t) \subset \pi^{-1}(g(z, t))$. Then we have $\tau h(z, t) \subset \tau\pi^{-1}(g(z, t)) = G(g(z, t))$, so that $\tau h(z, t)$ is a homotopy of $f(z)$ in U which covers the homotopy $g(z, t)$ in V.

[*Note added in proof* (June 6, 1952). Professor Beno Eckmann has recently called my attention to the fact that essentially the same concept as that of a fibre system has been introduced by him under the name "retrahierbare Ueberdeckung" in his paper "Zur Homotopietheorie gefaserter Raeume," *Commentarii Mathematici Helvetici*, vol. 14 (1941), pp. 141-192. In the first part of that paper the covering homotopy theorem was proved for a "retrahierbare Ueberdeckung", and an application was made of this theorem to the covering of a sphere by its great spheres.]

Let U be a non-singular algebraic variety of dimension r, and let Φ be an irreducible algebraic correspondence of dimension t between U and an algebraic variety V of dimension s. Then for a generic point η in V the set $\Phi^{-1}(\eta)$ is an irreducible algebraic variety of dimension $d = t - s$ in U, and we can consider $\Phi^{-1}(\eta)$ as a generic element of an algebraic system of d-cycles in U. For any point y in V the variety $\Phi^{-1}(y)$ is the carrier of the set of all d-cycles which are specializations of the d-cycle $\Phi^{-1}(\eta)$ over the specialization $\eta \to y$. A point y in V is said to be semiregular with respect to the correspondence Φ (or rather the inverse correspondence Φ^{-1}), if there is a uniquely determined specialization cycle of $\Phi^{-1}(\eta)$ over the specialization $\eta \to y$, and if this specialization cycle has no multiple components. It is easily seen that a point y in V is semi-regular with respect to Φ if and only if the variety $\Phi^{-1}(y)$ has the dimension d and the same degree (in the ambient projective space) as the variety $\Phi^{-1}(\eta)$, so that we can consider $\Phi^{-1}(y)$ itself as the specialization cycle of $\Phi^{-1}(\eta)$ over the specialization $\eta \to y$.

Since U is a differentiable manifold (of class C^∞), we can introduce in U a Riemannian metric. In fact, let $M = \{M_i\}$ be a locally finite system of

coordinate neighborhood covering U, and let the set of differentiable functions $\{e_i(x)\}$ be a partition of unity subordinate to this covering; then, if we denote by ds_i^2 the Euclidean metric of the coordinate neighborhood M_i, the differential form $ds^2 = \sum_i e_i(x)\, ds_i^2$ defines a Riemannian metric on U. We observe that by means of a suitable choice of the covering neighborhood M_i and the partition functions $e_i(x)$ we can make the Riemannian metric ds^2 in a sufficiently small neighborhood of any given point equal to the Euclidean metric with respect to any given coordinate system around this point. This fact will be convenient for us later. Let y' be a semi-regular point in V, and let R be a compact subset in $\Phi^{-1}(y')$ consisting of only simple points in $\Phi^{-1}(y')$. If $p(x)$ is any differentiable function which assigns to each point x in R a $(2r-2d)$-dimensional direction element which is transversal to the tangent space of $\Phi^{-1}(y')$ at the point x, then there exists a differentiable system of $(2r-2d)$-dimensional geodesic surfaces (or geodesic $(2r-2d)$-surfaces) $P(x)$ $(x \subset R)$, such that for each point x in R the surface $P(x)$ has the tangential direction $p(x)$ at x. If N is a sufficiently small neighborhood of y' in V, then for every point y in N and every point x in R, the intersection $\Phi^{-1}(y) \cap P(x)$ consists of exactly one point which is simple in $\Phi^{-1}(y)$, and the mapping $x \rightarrow \Phi^{-1}(y) \cap P(x)$ is a homeomorphism of R onto a compact subset $R(y)$ in $\Phi^{-1}(y)$. Thus each point x in $\sum_{y \subset N} R(y)$ is contained in exactly one geodesic $(2r-2d)$-surface of the system, which we shall also denote by $P(x)$; the function $\phi_N(x, y) = P(x) \cap R(y)$ $(x \subset \sum_{y \subset N} R(y), y \subset N)$ is then a slicing function for the system $R(y)$ $(y \subset N)$.

In case the variety $\Phi^{-1}(y')$ is non-singular, we can set $R = \Phi^{-1}(y')$ and hence $R(y) = \Phi^{-1}(y)$ for all y in N, so that the system of varieties $\Phi^{-1}(y)$ $(y \subset N)$ defines a fibre system in U. Now, if the generic variety $\Phi^{-1}(\eta)$ *is* non-singular, then there exists a proper subvariety H in V such that every point y in $V - H$ is semi-regular with respect to Φ and the variety $\Phi^{-1}(y)$ is non-singular. It follows then that the system of varieties $\Phi^{-1}(y)$ $(y \subset V - H)$ defines a fibre system in U. We shall now show that in case $d = 1$ this assertion is also true (with a suitable definition of H) even if $\Phi^{-1}(\eta)$ has singular points. We shall say that a point y' in V is regular with respect to Φ, if y' is semi-regular and if the variety $\Phi^{-1}(y')$ has the same singularities as the generic variety $\Phi^{-1}(\eta)$, i. e. each singular branch of the curve $\Phi^{-1}(y')$ is the specialization of a singular branch of the same order of the curve $\Phi^{-1}(\eta)$. It is easily seen that the set of all points in V which are not regular with respect to Φ is a proper subvariety in V, which we shall also denote by H. Let y' be a point in $V - H$, and let x' be a singular point in $\Phi^{-1}(y')$; then

there is a coordinate neighborhood M of x' in U and a suitably chosen system of coordinates $u_1, \cdots u_r$ in M with origin at x', such that for every point y in a sufficiently small neighborhood N of y' in V, the curve $\Phi^{-1}(y) \cap M$ is a regular analytic covering space of a fixed degree g over a neighborhood M_1 of the origin in the complex u_1-plane with a unique branch point over the origin $u_1 = 0$. Let M'_1 be a circular region $|u_1| < \epsilon$ $(\epsilon > 0)$ such that its closure $\bar{M}'_1$ is contained in M_1, and let $P(\alpha)$ $(\alpha \subset M_1)$ be the "hyperplane" in M defined by the equation $u_1 = \alpha$. Then, for each $u_1 \neq 0$ in $\bar{M}'_1$, there exist g disjoint circular domains $P_i(u_1)$, $i = 1, \cdots, g$, in $P(u_1)$, such that for each y in N and for each $i = 1, \cdots, g$, the intersection $\Phi^{-1}(y) \cap P_i(u_1)$ consists of exactly one point, while for $u_1 = 0$, the intersection $\Phi^{-1}(y) \cap P(0)$ itself consists of exactly one point, namely the branch point of $\Phi^{-1}(y)$ over $u_1 = 0$. For each y in N, we set $L(y) = \sum_{u_1 \subset M'_1} \Phi^{-1}(y) \cap P(u_1)$ and $\bar{L}(y) = \sum_{u_1 \subset \bar{M}'_1} \Phi^{-1}(y) \cap P(u_1)$ so that $\bar{L}(y)$ is the closure of the domain $L(y)$ in $\Phi^{-1}(y)$. We can then define a slicing function for the system $\bar{L}(y)$ $(y \subset N)$ by setting $\phi_N(x, y) = \Phi^{-1}(y) \cap P_i(u_1)$, $(x \subset \sum_{y \subset N} \bar{L}(y), y \subset N)$, where $P_i(u_1)$ is the one of the g domains in $P(u_1)$ which contains the point x. Furthermore, if we choose our Riemannian metric in U in such a way that it coincides with the Euclidean metric in the coordinate neighborhood M, then each $P_i(u_1)$ is a geodesic $(2r-2)$-surface, and for each point $x \neq x'$ in $\bar{L}(y')$ the $P_i(u_1)$ passing through it has a tangential direction element $p(x)$ which is transversal to the tangential space of $\Phi^{-1}(y')$ at x. In particular this function $p(x)$ is defined on the boundary $\bar{L}(y') - L(y')$ curve of the domain $L(y')$, and it is also differentiable. Let now $x^{(i)}$, $i = 1, \cdots, a$, be the singular points of $\Phi^{-1}(y')$, and let $M^{(i)}$, $i = 1, \cdots, a$, be suitably chosen (disjoint) coordinate neighborhoods of the points $x^{(i)}$, $i = 1, \cdots, a$, respectively in U, and let the Riemannian metric in U be so chosen that it coincides with the Euclidean metric in each $M^{(i)}$. If we choose the neighborhood N of y' sufficiently small, then we can define a fibre system $\bar{L}^{(i)}(y)$ $(y \subset N)$ and a slicing function $\phi^{(i)}{}_N(x, y)$ $(x \subset \sum_{y \subset N} \bar{L}^{(i)}(y), y \subset N)$ in each $M^{(i)}$, and we have also the function $p^{(i)}(x)$ defined in $\bar{L}^{(i)}(y') - L^{(i)}(y')$. If we set $Q(y) = \sum_{i=1}^{a} L^{(i)}(y)$ and $\bar{Q}(y) = \sum_{i=1}^{a} \bar{L}^{(i)}(y)$, then we can consider all the functions $\phi^{(i)}{}_N(x, y)$ together as a single slicing function $\phi_N(x, y)$ $(x \subset \sum_{y \subset N} \bar{Q}(y), y \subset N)$ for the fibre system $\bar{Q}(y)$ $(y \subset N)$, and also all the functions $p^{(i)}(x)$ together as a single function $p(x)$ defined on the boundary $\bar{Q}(y') - Q(y')$ of the domain $Q(y')$. For each $y \subset N$, let $R(y) = \Phi^{-1}(y) - Q(y)$; it is clear that $\bar{Q}(y)$

and $R(y)$ are two complementary closed domains in $\Phi^{-1}(y)$ with the curve $\bar{Q}(y) - Q(y)$ as their common boundary. In particular the boundary of $R(y')$ is $\bar{Q}(y') - Q(y')$, and the function $p(x)$ is defined and differentiable on the boundary $\bar{Q}(y') - Q(y')$. If we assign to each point x in $R(y')$ the set of all $(2r-2)$-dimensional direction elements at x which are transversal to the tangent space of $\Phi^{-1}(y')$ at x, then we obtain a differentiable fibre bundle over the space $R(y')$, in which the fibre is topologically a $(4r-4)$-cell. It follows then that the differentiable function $p(x)$, which is defined on the boundary of $R(y')$, can be extended to a differentiable function $p'(x)$ in the entire space $R(y')$. If we denote by $P'(x)$ $(x \subset R(y'))$ the system of geodesic $(2r-2)$-surfaces corresponding to the function $p'(x)$, then for each y in N, provided N is taken sufficiently small, the mapping $x \to \Phi^{-1}(y) \cap P'(x)$ is a homeomorphism of $R(y')$ onto a closed domain in $\Phi^{-1}(y)$ whose boundary is $\bar{Q}(y) - Q(y)$ and which approaches $R(y')$ as y approaches y'; hence this domain must be $R(y)$. If we denote by $P'(x)$, for any point x in $\sum_{y \subset N} R(y)$, the one geodesic $(2r-2)$-surface of this system which contains the point x, then the function

$$\phi'_N(x,y) = R(y) \cap P'(x) = \Phi^{-1}(y) \cap P'(x) \quad (x \subset \sum_{y \subset N} R(y),\, y \subset N)$$

is a slicing function of the system $R(y)$ $(y \subset N)$. Since, for

$$x \subset (\sum_{y \subset N} R(y)) \cap (\sum_{y \subset N} \bar{Q}(y)),\ y \subset N,$$

we have

$$\phi_N(x,y) = \Phi^{-1}(y) \cap P_i(u_1) = \Phi^{-1}(y) \cap P'(x) = \phi'_N(x,y),$$

the two slicing function $\phi_N(x,y)$ and $\phi'_N(x,y)$ are concordant whenever both are defined and hence can be considered together as one slicing function for the system of curves $\Phi^{-1}(y)$ $(y \subset N)$. Thus we have shown that the system of curves $\Phi^{-1}(y)$ $(y \subset V - H)$ is a fibre system.

2. THEOREM 1. *Let U be a non-singular algebraic variety of dimension r, and let a subvariety G of dimension d be a member of an irreducible algebraic system $G(y)$ $(y \subset V)$, with U as its carrier variety, which has at least one base point, and whose generic member is irreducible. If $f(z)$ is a continuous mapping of the unit interval I into U, with $f(0) = f(1) = x^{(0)}$ in G, then a finite power of this mapping is homotopic* rel. $z = 0, 1$, *to a continuous mapping of I into G. If the system $G(y)$ is involutional, then $f(z)$ itself is homotopic* rel. $z = 0, 1$, *to a continuous mapping of I into G.*

Remark. The m-th power of $f(z)$ is the mapping $\bar{f}(z)$ of I defined by setting $\bar{f}(z) = f(mz - i)$ for $i/m \leq z \leq (i+1)/m$, $i = 0, 1, \cdots, m-1$. The system $G(y)$ is said to be involutional, if it is induced by a rational transformation of U onto V.

Proof. Let the algebraic system be defined by an irreducible correspondence Φ between U and the variety V, and let $y^{(0)}$ be the point in V such that $\Phi^{-1}(y^{(0)}) = G(y^{(0)}) = G$. It is sufficient to prove our theorem for the case where $y^{(0)}$ is any point in an everywhere dense subset in V; this follows from the fact that for any point y in V the variety $\Phi^{-1}(y)$ is a neighborhood deformation retract in U. We begin with the special case where the system $G(y)$ is the linear system cut out on U by the system of linear subspaces of dimension $n - r + d$ $(d \geq 1)$ in the ambient space S_n, which all pass through a sufficiently general $S_{n-r+d-1}$. Since in this case $\Phi^{-1}(\eta)$ is non-singular for a generic point η in V, there exists a subvariety H in V such that each point y in $V - H$ is semi-regular and $\Phi^{-1}(y)$ is non-singular. We can assume without any loss of generality that $y^{(0)}$ is a point in $V - H$; furthermore, we can also assume that $x^{(0)}$ is a point in G outside of $\Phi^{-1}(H)$. Since $\Phi^{-1}(H)$ is a proper subvariety in U and hence topologically a subcomplex of dimension $\leq 2r - 2$ in the $2r$-dimensional topological manifold U, we can assume that, after a suitable homotopy (rel. $z = 0, 1$) if necessary, $f(z)$ is a mapping of I into $U - \Phi^{-1}(H)$. Then the mapping $g(z) = \Phi(f(z))$ of I into $V - H$ is well defined, and we have evidently $g(0) = g(1) = y^{(0)}$ and $f(z) \subset G(g(z))$ for all z. Let $x^{(1)}$ be any point in $G \cap S_{n-r+d-1}$, and let $h(z)$ be any continuous mapping of I into G such that $h(0) = x^{(0)}$ and $h(1) = x^{(1)}$. We set

$$f'(z) = \begin{cases} f(4z) & (0 \leq z \leq \tfrac{1}{4}), \\ h(4z-1) & (\tfrac{1}{4} \leq z \leq \tfrac{1}{2}), \\ x^{(1)} & (\tfrac{1}{2} \leq z \leq \tfrac{3}{4}), \\ h(4-4z) & (\tfrac{3}{4} \leq z \leq 1), \end{cases}$$

and

$$g'(z) = \begin{cases} g(4z) & (0 \leq z \leq \tfrac{1}{4}), \\ y^{(0)} & (\tfrac{1}{4} \leq z \leq \tfrac{1}{2}), \\ g(3-4z) & (\tfrac{1}{2} \leq z \leq \tfrac{3}{4}), \\ y^{(0)} & (\tfrac{3}{4} \leq z \leq 1); \end{cases}$$

it is clear that $f'(z) \simeq f(z)$ rel. $z = 0, 1$, and $g'(z) \simeq 0$ rel. $z = 0, 1$. Since $f'(z) \subset G(g'(z))$ for all z, it follows from the covering homotopy theorem that there is a homotopy of $f'(z)$ rel. $z = 0, 1$, which deforms $f'(z)$ into a mapping of I into G; hence $f(z)$ is also homotopic rel. $z = 0, 1$, to a mapping of I into G.

Turning to the general case, we observe first that we can assume without any loss of generality the following: (1) For a generic point ξ in U the variety $\Phi(\xi)$ consists of a finite number m of points; for otherwise we can replace V by its intersection with a suitably chosen linear subspace (passing through $y^{(0)}$) in its ambient space. It is clear that we have $m=1$ in case $G(y)$ is an involutional system. (2) For a generic point η in V the variety $\Phi^{-1}(\eta)$ is a curve, i. e. $d=1$; for otherwise we can replace U by its intersection with a suitably chosen S_{n-d+1} (passing through a base point of the system $G(y)$) in S_n, and we have just shown above that any continuous mapping $f(z)$ of I into U, with $f(0)=f(1)$ in $U \cap S_{n-d+1}$, is homotopic rel. $z=0,1$, to a continuous mapping of I into $U \cap S_{n-d+1}$. Now, let H be the subvariety in V containing all points which are not regular with respect to Φ, so that the system of curves $G(y)$ $(y \subset V-H)$ is a fibre system; let T be the subvariety in U such that for every point x in $U-T$ the set $\Phi(x)$ consists of m distinct points outside of H. Without any loss of generality we can assume that $y^{(0)}$ is a point in $V-H$ such that $G=\Phi^{-1}(y^{(0)})$ is not entirely in T, and that $x^{(0)}$ is a point in $G-T$; then $\Phi(x^{(0)})$ consists of m points, one of which is the point $y^{(0)}$. Since $T+\Phi^{-1}(H)$ is a proper subvariety in U and hence topologically a subcomplex of dimension $\leq 2r-2$ in the $2r$-dimensional manifold U, we can assume that, after a suitable homotopy (rel. $z=0,1$) if necessary, $f(z)$ is a mapping of I into $U-T-\Phi^{-1}(H)$. Then the image $\Phi(f(z))$ consists of m distinct mappings of I into $V-H$, one (and only one) of which will be a mapping $g(z)$ such that $g(0)=y^{(0)}$. The point $g(1)$ is one of the m points in the set $\Phi(x^{(0)})$, though not necessarily the point $y^{(0)}$. It is easily seen that if $\bar{f}(z)$ is the m-th power of $f(z)$, then the image $\Phi(\bar{f}(z))$ of $\bar{f}(z)$ will consist of m distinct mappings of I into $V-H$, one of which will be a mapping $\bar{g}(z)$ such that $\bar{g}(0)=\bar{g}(1)=y^{(0)}$. Let $x^{(1)}$ be a base point of the system $G(y)$, and let $h(z)$ be a continuous mapping of I into G such that $h(0)=x^{(0)}$ and $h(1)=x^{(1)}$. If we now define $f'(z)$ and $g'(z)$ again as before, replacing $f(z)$ and $g(z)$ by $\bar{f}(z)$ and $\bar{g}(z)$ respectively, we can repeat exactly the same argument and conclude that $\bar{f}(z)$ is homotopic rel. $z=0,1$, to a mapping of I into G. This concludes the proof of Theorem 1.

In the following we shall denote by $F(U)$ the fundamental group of a topological space U, considered as a group of mapping classes with some one fixed reference point. If W and X are two subsets in U, then the identity mapping of $W \cap X$ into W will induce a homomorphism of $F(W \cap X)$ into $F(W)$, the reference point in both groups being one and the same point in $W \cap X$; we shall then denote by $F(W, X)$ the subgroup of $F(W)$ which is the image of $F(W \cap X)$ under this homomorphism.

THEOREM 2. *Let Φ be a rational transformation of a non-singular algebraic variety U of dimension r onto a non-singular algebraic variety V of dimension s, with the properties*: (1) *For a generic point η in V the variety $\Phi^{-1}(\eta)$ is irreducible, and* (2) *the set H of all points which are not semi-regular with respect to Φ is a subvariety of dimension $\leq s-2$ in V. Then there is a homomorphism of $F(U)$ onto $F(V)$, and the kernel of this homomorphism is the subgroup $F(U, \Phi^{-1}(y^{(0)}))$, where $y^{(0)}$ is any sufficiently general point in V.*

Proof. Let T be the fundamental variety of Φ in U; it is well known that T has a dimension $\leq r-2$. It is clear that the variety $\Phi^{-1}(\eta)$ has the dimension $d = r-s$; let c be the degree of $\Phi^{-1}(\eta)$ considered as a variety in the ambient space S_n of U. Let $y^{(0)}$ be a point in V such that $\Phi^{-1}(y^{(0)})$ is not contained in T, and let $x^{(0)}$ be a point in $\Phi^{-1}(y^{(0)})-T$. We shall take $x^{(0)}$ as the reference point of the groups $F(U)$, $F(U-T)$, $F(\Phi^{-1}(y^{(0)}))$, $F(U, \Phi^{-1}(y^{(0)}))$, and $F(U-T, \Phi^{-1}(y^{(0)}))$, and take $y^{(0)}$ as the reference point of the group $F(V)$. The identity mapping of $U-T$ into U induces a homomorphism θ of $F(U-T)$ into $F(U)$; since T is topologically a subcomplex of dimension $\leq 2r-4$ in the $2r$-dimensional manifold U, it is easily seen that θ is an isomorphism of $F(U-T)$ onto $F(U)$, and that the image of $F(U-T, \Phi^{-1}(y^{(0)}))$ under θ is precisely $F(U, \Phi^{-1}(y^{(0)}))$. Since Φ is a continuous mapping of $U-T$ into V, there is an induced homomorphism ϕ of $F(U-T)$ into $F(V)$. We set $\Delta = \phi\theta^{-1}$, so that Δ is a homomorphism of $F(U)$ into $F(V)$. We shall prove our theorem by showing that (provided the point $y^{(0)}$ is sufficiently general) Δ is a homomorphism of $F(U)$ onto $F(V)$ and its kernel is $F(U, \Phi^{-1}(y^{(0)}))$.

Let L be a linear subspace of dimension $n-r+s$ in the ambient space S_n of U, such that $U \cap L$ is a non-singular variety of dimension s. The rational transformation Φ will then induce a rational transformation $\bar{\Phi}$ of $U \cap L$ onto V, and we have $\bar{\Phi}^{-1}(y) \subset \Phi^{-1}(y) \cap L$ for every point y in V. Since, for a general point η in V, the set $\bar{\Phi}^{-1}(\eta) = \Phi^{-1}(\eta) \cap L$ consists of c distinct points in $U-T$, there exists a proper subvariety K in V such that for every point y in $V-K$ the set $\bar{\Phi}^{-1}(y)$ consists of c distinct points in $U-T$ (which are then necessarily simple points in the variety $\Phi^{-1}(y)$). We shall also assume that $y^{(0)}$ is a point in $V-K$ and $x^{(0)}$ is a point in $\bar{\Phi}^{-1}(y^{(0)})$.

Let $g(z)$ be a continuous mapping of the unit interval I into V such that $g(0) = g(1) = y^{(0)}$. Since K is topologically a subcomplex of dimension $\leq 2s-2$ in the topological manifold V of dimension $2s$, we can assume, after

a homotopy (rel. $z=0,1$) if necessary, that $g(z)$ is a mapping of I into $V-K$. Since $\bar{\Phi}^{-1}(V-K)$ is a c-fold regular covering space of $V-K$, the inverse image $\bar{\Phi}^{-1}(g(z))$ consists of c distinct mappings of I into $U-T$, one of which will be a mapping $f(z)$ such that $f(0)=x^{(0)}$. Let $x^{(1)}=f(1)$, and let $f_1(z)$ be any continuous mapping of I into $\Phi^{-1}(y^{(0)})-T$ such that $f_1(0)=x^{(1)}$ and $f_1(1)=x^{(0)}$. If we set

$$f'(z)=\begin{cases} f(2z) & (0\leq z\leq \frac{1}{2}),\\ f_1(2z-1) & (\frac{1}{2}\leq z\leq 1),\end{cases}$$

then we have evidently a continuous mapping of I into $U-T$ such that $\Phi(f'(z))\simeq g(z)$ rel. $z=0,1$. This shows that Δ is a homomorphism of $F(U)$ onto $F(V)$.

Now let $f(z)$ be a continuous mapping of I into U such that $f(0)=f(1)=x^{(0)}$. According to Theorem 1, $f(z)$ is homotopic rel. $z=0,1$ to a mapping of I into $U\cap L$, and hence also homotopic rel. $z=0,1$ to a mapping of I into $\bar{\Phi}^{-1}(V-K)$. We can therefore assume that $f(z)$ is already a mapping of I into $\bar{\Phi}^{-1}(V-K)$; then the mapping $g(z)=\Phi(f(z))$ is defined, and we have $g(0)=g(1)=y^{(0)}$. In order to prove that the kernel of Δ is $F(U,\Phi^{-1}(y^{(0)}))$, we have only to show that if the mapping $g(z)$ is homotopic to zero rel. $z=0,1$ in V, then $f(z)$ is homotopic rel. $z=0,1$ to a mapping of I into $\Phi^{-1}(y^{(0)})$. Without loss of generality we can restrict ourselves to the case $d=1$; for otherwise we can replace U by its intersection with a suitably chosen linear subspace of dimension $n-r+s+1$ which contains the space L. It is then easily seen that we can assume the subvariety K so chosen that every point in $V-K$ is regular with respect to Φ; then the system of curves $\Phi^{-1}(y)$ $(y\subset V-K)$ is a fibre system. Furthermore, there exists a subvariety H' of dimension $\leq s-2$ in V, which contains H and is contained in K, such that for every point y in $V-H'$ the set $\bar{\Phi}^{-1}(y)$ consists of at most c points, each of which is either a simple point in $\Phi^{-1}(y)$ or a singular point of a "generic" nature.

Since K is topologically a subcomplex of dimension $\leq 2s-2$, and H' is topologically a subcomplex of dimension $\leq 2s-4$ in the $2s$-dimensional manifold V, there exists in $V-H'$ a homotopy $g(z,t)$ of the mapping $g(z)$ rel. $z=0,1$ with the following properties: (1) $g(z,t)\subset V-K$ for all z and t except a finite number m of points $g((2i-1)/2m,1)$, $i=1,\cdots,m$, which all belongs to $K-H'$; (2) $g(i/m,1)=y^{(0)}$ for $i=1,\cdots,m$, and $g(z,1)=g((2i-1-mz)/m,1)$ for $(i-1)/m\leq z\leq i/m$. Since $\bar{\Phi}^{-1}(V-H')$ is an analytic covering space (with branch points) of $V-H'$, and since moreover $\bar{\Phi}^{-1}(V-K)$ is a regular covering space of $V-K$, there corre-

sponds to the homotopy $g(z,t)$ a uniquely determined homotopy $f(z,t)$ rel. $z = 0,1$ in $\bar{\Phi}^{-1}(V-H')$, with $f(z,0) = f(z)$ and $f(z,t) \subset \bar{\Phi}^{-1}(g(z,t))$. Therefore, to prove our assertion, it is sufficient to consider the following problem. Let $g(z)$ be a continuous mapping of I into $V-H'$ with the properties: $g(0) = g(1) = y^{(0)}$; $g(z) = g(1-z)$ for all z; $g(z) \subset V-K$ for all $z \neq \frac{1}{2}$, and $g(\frac{1}{2})$ belongs to $K-H'$. Let $f(z)$ be a continuous mapping of I into $\bar{\Phi}^{-1}(V-H')$ such that $f(z) \subset \bar{\Phi}^{-1}(g(z))$ for all z; then we have to show that $f(z)$ is homotopic rel. $z = 0,1$ in U to a mapping of I into $\Phi^{-1}(y^{(0)})$.

Let $x^{(1)} = f(\frac{1}{2})$ and $y^{(1)} = g(\frac{1}{2})$, and let R be a compact neighborhood of $x^{(1)}$ in $\Phi^{-1}(y^{(1)})$ which contains no singular points of $\Phi^{-1}(y^{(1)})$ except possibly the point $x^{(1)}$ itself. Then, as we have shown in the preceding section, there exists a neighborhood N of $y^{(1)}$, and to every point y in N there corresponds a compact subset $R(y)$ in $\Phi^{-1}(y)$ (with $R(y^{(1)}) = R$), such that the system $R(y)$ $(y \subset N)$ is a fibre system; furthermore, there is a neighborhood M of $x^{(1)}$ in U such that $M \cap \Phi^{-1}(y) \subset R(y)$ for all y in N. Let I_e $(e > 0)$ denote the interval $|z-\frac{1}{2}| \leq e$; and let e be taken so small that $f(z) \subset M$ and $g(z) \subset N$ for all z in I_e. Since the mapping $g(z)$ of I_e into N is evidently homotopic rel. $z = \frac{1}{2}-e, \frac{1}{2}+e$ in N to a mapping of I_e into the point $y^{(2)} = g(\frac{1}{2}-e) = g(\frac{1}{2}+e)$, it follows from the covering homotopy theorem that the mapping $f(z)$ of I_e into $M \cap \Phi^{-1}(N)$ is also homotopic rel. $z = \frac{1}{2}-e, \frac{1}{2}+e$ to a mapping $h(z)$ of I_e into $\Phi^{-1}(y^{(2)})$. If we set

$$f'(z) = \begin{cases} f(z) & (|z-\frac{1}{2}| \geq e), \\ h(z) & (|z-\frac{1}{2}| \leq e), \end{cases}$$

$$g'(z) = \begin{cases} g(z) & (|z-\frac{1}{2}| \geq e), \\ y^{(2)} & (|z-\frac{1}{2}| \leq e), \end{cases}$$

then we have evidently $f(z) \simeq f'(z)$ rel. $z = 0,1$ in U and $g'(z) \simeq 0$ rel. $z = 0,1$ in $V-K$. Since $f'(z) \subset \Phi^{-1}(g'(z))$ for all z, it follows then from the covering homotopy theorem that $f'(z)$ is homotopic rel. $z = 0,1$ to a mapping of I into $\Phi^{-1}(y^{(0)})$, and hence $f(z)$ is also homotopic rel. $z = 0,1$ to a mapping of I into $\Phi^{-1}(y^{(0)})$. This concludes the proof of Theorem 2.

THE JOHNS HOPKINS UNIVERSITY.

INDUCED REPRESENTATIONS.*

By F. I. Mautner.

1. Introduction; statement of Theorem 1. Let G be a locally compact topological group and K an arbitrary (but fixed) compact subgroup of G. With every continuous unitary representation u of K in a Hilbert (or finite dimensional) space $\mathfrak{h}$ over the complex numbers we can associate a continuous unitary representation U of G, the so called *induced representation*, as follows:

Consider all those Haar-measurable functions $X(g)$ defined on G with values in $\mathfrak{h}$ for which [1]

$$\int_G \| X(g) \|^2 dg < \infty,$$

where $\| X(g) \|$ denotes, for fixed g, the norm of $X(g)$ as an element of the Hilbert space $\mathfrak{h}$. We restrict ourselves to only those functions $X(g)$ which satisfy also

$$X(gk) = u(k^{-1})X(g) \quad \text{for all } k \in K. \tag{1.1}$$

Clearly all such functions form a linear space over the complex numbers from which we obtain a Hilbert (or finite dimensional) space $\mathfrak{H}$ if we identify functions $X(g)$ which differ on sets of Haar-measure zero, and define an inner product (X, Y) in $\mathfrak{H}$ by

$$(X, Y) = \int_G (X(g), Y(g))dg, \tag{1.2}$$

where $(X(g), Y(g))$ denotes for, fixed $g \in G$, the inner product in the space $\mathfrak{h}$. Now define for every $\gamma \in G$ a linear transformation $U(\gamma)$ by

$$(U(\gamma)X)(g) = X(\gamma^{-1}g). \tag{1.3}$$

* Received April 19, 1951.

[1] Introduce in $\mathfrak{h}$ an arbitrary complete orthonormal system, and denote by $X_j(g)$ the expansion coefficients of $X(g)$ with respect to it (for each fixed $g \in G$). We shall say that the vector-valued function $X(g)$ is Haar-measurable if each of the complex valued functions $X_j(g)$ is Haar-measurable. It is clear that this definition is independent of the particular complete orthonormal system in $\mathfrak{h}$, and that it follows that the inner product $(X(g), Y(g))$ of any two Haar-measurable vector valued functions $X(g)$ and $Y(g)$ is Haar-measurable.

Clearly $U(\gamma)$ is a unitary operator of the space $\mathfrak{H}$ onto itself, and the mapping $U:\gamma\rightarrow U(\gamma)$ defines a continuous unitary representation of G in the space $\mathfrak{H}$. We call U the *induced representation generated by u*, and write

$$(1.4)\qquad U=\operatorname{ind} u \quad\text{or}\quad U=\operatorname*{ind}_{K\uparrow G} u.$$

For finite groups this is the same as the classical definition of "induced representation." For an arbitrary locally compact group and a closed subgroup a definition has recently been given by Mackey [5]. It is easy to see that Mackey's definition reduces to the above in the case when the subgroup K is compact, which we shall assume throughout.

Suppose for the moment that G is compact too. Then the representation U of G can be decomposed into a discrete (i. e. ordinary) direct sum $\sum_{\oplus}\mu_j M_j$ of irreducible finite dimensional pairwise inequivalent representations M_j of G, each occurring with multiplicity μ_j. The classical Frobenius reciprocity theorem asserts in this case that

$$(1.5)\qquad \text{multiplicity of } u \text{ in } M_j(K)=\mu_j,$$

where $M_j(K)$ denotes the restriction of the representation M_j to the subgroup K.

The problem arises whether this theorem can be generalized to the case where G is no longer compact. It is known that U can still be decomposed into irreducible unitary representations M_j in the sense of generalized direct sums (= direct integrals). Therefore equation (1.5) can still be formulated. However [2] there exist infinite discrete groups for which (1.5) is false even when one takes for K the trivial subgroup = (1).

But one can replace (1.5) by the following formulation, which is equivalent to (1.5) in the classical case (G compact). Denote by U_j the repetition μ_j times of M_j:

$$U_j=M_j\oplus M_j\oplus M_j\oplus\cdots\oplus M_j.$$

Then we have $U=\sum_{\oplus} U_j$, where U_i and U_j are inequivalent for $i\neq j$. Denote by W_j the algebra generated by the operators $U_j(g)$ in the representation space $\mathfrak{H}_j$ of U_j, and by W'_j the commuting algebra of W_j in $\mathfrak{H}_j$. Then the classical theory of linear algebras tells us that

[2] G. W. Mackey has found an explicit decomposition of the regular representation of certain discrete groups for which every irreducible component representation occurs with multiplicity one, and is infinite dimensional (oral communication of an unpublished result).

$$\dim(W'_j) = \mu_j^2, \tag{1.6}$$

where $\dim(W'_j) =$ dimension of W'_j as a linear space over the complex numbers. Then we have

$$\text{multiplicity of } u \text{ in } U_j(K) = \dim(W'_j). \tag{1.7}$$

Moreover it is well known that if $E_1, E_2, \cdots$ denote the minimal self-adjoint idempotents in the center Z of W, then $W_j = E_jW$, and $W'_j = E_jW'$. This suggests another possibility of generalization, namely the use of von Neumann's central decomposition. And, in fact, form (1.7) of the Frobenius reciprocity theorem can be generalized as follows. Assume now that G is an arbitrary locally compact group whose Haar measure is both left and right invariant, and which satisfies the second aviom of countability. Then the above Hilbert space $\mathfrak{H}$ is separable, so we can apply Theorem VII of [15]. Indeed, let W be the weakly closed self-adjoint algebra of bounded linear operators generated by the operators $U(g)$ in $\mathfrak{H}$, W' the commuting algebra of W, and Z the center,, i. e. $Z = W \cap W'$. Then we obtain a direct integral decomposition

$$\mathfrak{H} = \int_{\oplus} \mathfrak{H}_t \tag{1.8}$$

under the operators $U(g)$ to which the center Z "belongs" in the sense af Theorems IV and VII of [15]. For each $g \varepsilon G$ we obtain an operator-valued function $U(g, t)$ of t which can be changed arbitrarily on t-sets of measure zero. It has been proved in Theorem 1.1 of [8] that one can find for each $g \varepsilon G$ one such operator-valued function $U(g, t)$, and for each t a continuous unitary repersentation $\mathfrak{V}_t$: $g \to \mathfrak{V}_t(g)$ of G in the space $\mathfrak{H}_t$ such that

$$U(g, t) = \mathfrak{V}_t(g) \text{ for } g \notin N_t, \tag{1.9}$$

where N_t is some subset of G of Haar measure zero, depending on t. Then we have

THEOREM 1. *Let G be a locally compact unimodular group, satisfying the second axiom of countability. Let K be an arbitrary compact subgroup of G and u: $k \to u(k)$ a continuous irreducible (unitary) representation of the subgroup K. Perform the central decomposition (1.8) of the space $\mathfrak{H}$ of the induced representation $U = \operatorname{ind} u$. Denote by $\mathfrak{V}_t(K)$ the restriction of the representation $\mathfrak{V}_t$ of G to the subgroup K, and by $[\mathfrak{V}_t(g)]'$ the algebra of all those bounded operators in $\mathfrak{H}_t$ which commute with $\mathfrak{V}_t(g)$ for every $g \varepsilon G$.*

ASSERTION.

(1.10) *Multiplicity of u in* $\mathfrak{V}_t(K) = \dim\{[\mathfrak{V}_t(g)]'\}$ *for almost every t,*

where "dim" denotes the ordinary dimension of a linear space over the complex numbers.

This theorem will be proved in **2** to **5**. Some of the lemmas proved in **2** to **5** may also be of interest independently of the proof of Theorem 1. There is some overlap between Theorem 1 above and the generalizations of the Frobenius reciprocity theorem recently obtained by Mackey [6]. The restrictions under which we prove Theorem 1 above are different from the assumption under which Mackey's results hold.

This suggests the possibility of a more general result which should contain all the known generalizations of the Frobenius reciprocity theorem. We shall not discuss this further generalization, which seems to present serious difficulties, in the present paper. However Theorem 1 in its present form has various applications. The simplest application is to the case where the commuting algebra W' of a given induced representation $\operatorname{ind} u$ is known to be commutative. For W' commutative implies $\dim\{[\mathfrak{V}_t(g)]'\} = 1$ for almost every t; hence by Theorem 1 the multiplicity of u in $\mathfrak{V}_t(K)$ equals 1 for almost every t. This result is the main step in the derivation of the Plancherel formula outlined in [11a]. The details of this derivation were included in the original version of this paper, but have been separated from the rest of the paper at the suggestion of the referee.

Theorem 1 can also be applied to the case where G is a semi-simple Lie Group, which for simplicity of statement we take to be its own adjoint group, and K a maximal compact subgroup of G. The results of Harish-Chandra together with Theorem 1 above imply in this case that the above commuting algebra W' is a finite module over its center for every irreducible continuous representation u of K. This implies that the methods and results outlined in [11a] apply to the representation space $\mathfrak{H}$ of $\operatorname{ind} u$ in this case for every u. Thus one obtains by very general considerations a rather special kind of Plancherel formula for each $\mathfrak{H}$, and hence also for the space $\mathfrak{L}_2(G)$ of all complex valued Haar-Lebesgue square integrable functions on G.

We shall use in this paper essentially the same terminology and notation as in [7] and [8]. In particular Hilbert spaces are always assumed to admit complex scalars and will be denoted by $\mathfrak{H}$, $\mathfrak{h}$ or $\mathfrak{H}_t$, and occasionally also by $\boldsymbol{H}$ or $\boldsymbol{H}_t$. Whenever we use the results of [15] in an essential manner the Hilbert spaces in question have to be assumed to be separable (they may be

finite dimensional). For the notions and properties of generalized direct sums $\mathfrak{H} = \int_{\oplus} \mathfrak{H}_t$ we refer the reader to [15]. If we are given such a generalized direct sum (= direct integral), then there corresponds to every $x \varepsilon \mathfrak{H}$ a vector valued function $x(t)$ such that the value of $x(t)$ is for a given t an element of $\mathfrak{H}_t$. We call $x(t)$ the "*component*" of x in the space $\mathfrak{H}_t$ and say also that "*x decomposes into $x(t)$*." To define the direct integrals the points t must form a measure space. Our assertions about a given direct integral will always be about "almost all t" or "almost all the spaces $\mathfrak{H}_t$" etc. by which will be meant "all t (or $\mathfrak{H}_t$ etc.) except for a set of elements t whose measure (with which the given direct integral is formed) is zero." I. e. the measure referred to in connection with a given direct integral will always be the particular measure used to define the given direct integral, even when not mentioned explicitly.

It has been well known for some time that it is possible to introduce a topology on a measure space. It seems unlikely that the introduction of such a topology into the measure space used for the given direct integral will make it possible to eliminate the "almost all" statement from most of the deeper results on generalized direct sums. Thus we shall not introduce the above mentioned topology in the present paper, but base our assertions and proofs about direct integrals on [7], [8] and [15]. It is however clear that in any particular case where a topology is really wanted, our results and methods can readily be translated. This remark applies in particular to the results outlined in [11a]. There it seems of interest to consider the measure space in question more closely. We plan to come back to this in a later publication, where it will be shown that the methods of the present paper and of [11a] lead to more precise results especially for semi-simple Lie groups and throw some new light also on the problem of eigenfunction expansions for certain partial differential equations (both of the ellipitic and hyperbolic type) especially when there is a continuous spectrum.

2. Isomorphisms of factors. In this section let Ξ be an arbitrary Hilbert space over the complex numbers, and $\boldsymbol{M}$ a factor in Ξ; i. e. $\boldsymbol{M}$ is a weakly closed self-adjoint algebra of bounded operators in Ξ whose center consists of the scalar multiples of the identity operator I. According to [12] there exists on $\boldsymbol{M}$ an essentially unique relative dimension function $d(E)$ defined for all projections $E \varepsilon \boldsymbol{M}$. According to a result of Rickart (Cor. 4.13 of [17]) the projections $E \varepsilon \boldsymbol{M}$ with $d(E) < \infty$ are contained in every proper two-sided ideal of $\boldsymbol{M}$. There exists therefore a two-sided ideal $\boldsymbol{J}$ of $\boldsymbol{M}$ con-

taining all projections of finite relative dimension such that every other non-zero two-sided ideal of $\mathbf{M}$ contains $\mathbf{J}$. The following lemma is only stated explicitly for the convenience of the reader and is not essentially new.

LEMMA 2.1. *$\mathbf{M}$ is the smallest weakly closed self-adjoint operator algebra which contains the identity operator I and the ideal $\mathbf{J}$.*

Proof. If $\mathbf{M}$ is of finite type in the sense of [12], then it has no proper two-sided ideals; hence $\mathbf{J} = \mathbf{M}$ in this case, and hence our Lemma 2.1 is (trivially) true in this case.

If $\mathbf{M}$ is a factor of infinite type, then Theorem VIII of [12] implies that there exists for every integer $n \geq 0$ a projection $E_n \varepsilon \mathbf{M}$ with $d(E_n) = n$. It follows from Lemma 8.13 of [12] that we can assume $E_n < E_{n+1}$. Therefore the E_n converge strongly to a projection $E \varepsilon \mathbf{M}$. Since $d(E - E_n) \geq 0$, we have $d(E) = d(E - E_n) + d(E_n) \geq n$ for all n; thus $d(E) = \infty$. Now let F be an arbitrary projection $\mathbf{M}$ with $d(F) = \infty$. By Lemma 8.13 of [12] E and F are equivalent in the sense that there exists a partially isometric operator $\pi \varepsilon \mathbf{M}$ such that $E = \pi\pi^*$ and $F = \pi^*\pi$, whence $\pi^* E \pi = F$. Put $F_n = \pi^* E_n \pi$. Then strong $\lim_n F_n = F$ and $d(F_n) = d(F_n)$. So we have found for every projection $F \varepsilon \mathbf{M}$ with $d(F) = \infty$ and ascending sequence of projections $\varepsilon\, \mathbf{J}$ converging strongly to F. This together with the fact that $\mathbf{M}$ is generated (as a weakly or strongly closed operator algebra) by its projection (cf. [16]) proves Lemma 2.1.

As an immediate consequence we obtain

LEMMA 2.2. *Let $\mathbf{M}$ be a factor and C an arbitrary idempotent element of $\mathbf{M}'$, i. e. $C^2 = C$ is a bounded linear operator, and $CA = AC$ for all $A \varepsilon \mathbf{M}$. Then the mapping $A \to CA$ is an isomorphism of $\mathbf{M}$ onto the algebra $C\mathbf{M}$, provided $C \neq 0$.*

Proof. The mapping $A \to CA$ is clearly a homomorphism of the algebra $\mathbf{M}$ since $C^2 = C$ and $AC = CA$. Hence the set of elements A of $\mathbf{M}$ for which $AC = 0$ is a two-sided ideal of $\mathbf{M}$. If it were not the zero ideal it would have to contain $\mathbf{J}$, by the above mentioned result of Rickart. But then Lemma 2.1 above would imply $AC = 0$ for all $A \varepsilon \mathbf{M}$, which contradicts $IC = C \neq 0$, since the identity operator I is an element of $\mathbf{M}$.

3. Decomposition of an invariant subspace. In this section let $\mathbf{H}$ be an arbitrary (separable) Hilbert space and $\mathbf{W}$ an arbitrary weakly closed self-adjoint algebra of bounded operators in $\mathbf{H}$. Perform the central decomposition (Theorem VII of [15]) under the algebra $\mathbf{W}$:

$$H = \int_{\oplus} H_t. \tag{3.1}$$

Now let $E = E^* = E^2 \varepsilon W$, where W' denotes as usual the commuting algebra of W. By Theorem V of [15] $E \varepsilon W'$ implies that E is decomposable under (3.1) into an operator-valued function, say $E(t)$. Let T_1 be the set of those t for which $E(t) \neq 0(t)$, where $0(t)$ denotes the zero-operator in the space H_t. Put $EH = H_1$, and $E(t)H_t = H_{1,t}$, for $t \varepsilon T_1$. Clearly for any $x \varepsilon H$ we have $x \varepsilon H_1$ if and only if

$$\begin{aligned} x(t) &= 0(t) \quad \text{for } t \notin T_1 \\ \text{and} \quad x(t) &\varepsilon H_{1t} \quad \text{for } t \varepsilon T_1 \end{aligned} \tag{3.2}$$

(after a possible change of $x(t)$ on a t-set of measure zero which we assume to have been made). We have therefore a one-one correspondence between the elements x of H_1 and those equivalence classes of vector valued functions $x(t)$ which occur in (3.1) and satisfy the conditions (3.2). It is easy to conclude from this that this one-one correspondence defines a direct integral decomposition

$$H_1 = \int_{\oplus} H_{1t}, \tag{3.3}$$

where t now ranges only over the set T_1 and the measure used in the definition of (3.3) is the restriction of the measure used in the definition of (3.1) to the subset T_1.

Denote by Z the center of W $(Z = W \cap W')$, by W_1 the algebra EW, and by Z_1 the algebra EZ, where the elements of W_1 and Z_1 are considered to be operators of the space H_1 into itself. We then have

LEMMA 3.1. *The direct integral* (3.3) *is the central decomposition of the space H_1 under the algebra W_1 i. e. the algebra Z_1 "belongs to it" in the sense of* [15]. *In particular Z_1 is the center of W_1.*

Proof. Let $A_1 \varepsilon W_1$; then there exists an element A of W such that A_1 is the restriction of EA to H_1. Since $A \varepsilon W$, it is decomposable under (3.1) into an operator-valued function $A(t)$ say. Moreover $(EA)(t) = E(t)A(t)$ for almost all t implies

$$E(t)A(t)H_{1t} \subseteq H_{1t} \tag{3.4}$$

for almost all t. Now change $A(t)$ on a set of measure zero so that (3.4) becomes valid for *all* t and put

$$A_1(t) = \text{restriction of } E(t)A(t) \text{ to } H_{1t} \text{ for } t \varepsilon T_1. \tag{3.5}$$

Then A_1 decomposes into $A_1(t)$ under (3.3).

Let us now choose countably many elements $A^{(j)}$ $(j=1,2,\cdots)$ of $\boldsymbol{W}$ which generate $\boldsymbol{W}$ (in the weak or strong topology). It follows from the proof of Theorem VI (on p. 459) of [15] that we may assume $\boldsymbol{W}(t)$ to be generated by the elements $A^{(j)}(t)$ for every t where we can also assume that the $A^{(j)}(t)$ are chosen such that (3.4) holds for every t and also

$$(3.6) \qquad A^{(j)}(t)E(t)=E(t)A^{(j)}(t) \text{ for every } t \text{ and all } j=1,2,3,\cdots.$$

Now let $\boldsymbol{W}_1(t)$ be the (weakly closed self-adjoint) algebra generated by the operators $A_1^{(j)}(t)$ for $t \varepsilon \boldsymbol{T}_1$. Then it is clear that since $E(t) \varepsilon \boldsymbol{W}'(t)=\boldsymbol{W}(t)'$ (cf. Lemma 13 of [15]), the mapping

$$(3.7) \qquad A(t) \to A_1(t) = \text{restriction of } E(t)A(t) \text{ to } \boldsymbol{H}_{1t}$$

is a homomorphism of $\boldsymbol{W}(t)$ onto $\boldsymbol{W}_1(t)$ for almost every $t \varepsilon \boldsymbol{T}_1$. Since $E(t) \neq 0(t)$ for $t \varepsilon \boldsymbol{T}_1$, and $\boldsymbol{W}(t)$ is a factor for every t, we may apply Lemma 2.2 above and conclude that *the mapping* (3.7) *is an isomorphism of* $W(t)$ *onto* $W_1(t)$ *for almost every* $t \varepsilon \boldsymbol{T}_1$. Hence in particular $\boldsymbol{W}_1(t)$ is also a factor for almost every $t \varepsilon \boldsymbol{T}_1$.

Note that it follows from the way $\boldsymbol{W}_1(t)$ is defined that if X is an arbitrary element of $\boldsymbol{W}_1$, i. e. $X=AE=EA$ with $A \varepsilon \boldsymbol{W}$, then $X(t)=$ above $A_1(t)$ is an element of $\boldsymbol{W}_1(t)$ for almost every $t \varepsilon \boldsymbol{T}_1$. Observe also that $\boldsymbol{W}_1(t)$ depends measurably on t for $t \varepsilon \boldsymbol{T}_1$ in the sense of definition 5 of [15] under the direct integral (3.3). Hence in order to prove that (3.3) is the central decomposition under $\boldsymbol{W}_1$ it is by Theorem VII of [15] sufficient to prove the following: If $Y(t)$ is an arbitrary bounded measurable operator valued function (in the sense of Definition 5 of [15]) defined for all $t \varepsilon \boldsymbol{T}_1$, and satisfies $Y(t) \varepsilon \boldsymbol{W}_1(t)$, then there exists an operator Y of the space $\boldsymbol{H}_1$ which is an element of $\boldsymbol{W}_1$ and decomposes into $Y(t)$ under (3.3); i. e. we have to prove $\boldsymbol{W}_1 \approx \sum \boldsymbol{W}_1(t)$ in the terminology of [15] under the direct integral (3.3).

By Lemma 12 of [15] the ring generated by $\boldsymbol{P}_1$ and the operators $A_1^{(j)}$ satisfies $\approx \sum \boldsymbol{W}_1(t)$, where $\boldsymbol{P}_1$ denotes the ring of those operators of $\boldsymbol{H}_1$ which decomposes into scalars under (3.3). Since the $A_1^{(j)}$ generate $\boldsymbol{W}_1$ it remains to prove $\boldsymbol{P}_1 \subseteqq \boldsymbol{W}_1$.

$C_1 \varepsilon \boldsymbol{P}_1$ means that there exists a complex valued bounded measurable function $c(t)$ defined for $t \varepsilon \boldsymbol{T}_1$, such that C_1 decomposes under (3.3) into $C_1(t)=c(t)I_1(t)$, where $I_1(t)$ denotes the identity operator in $\boldsymbol{H}_{1t}$. Let $C(t)$ denote $c(t)I(t)$ or $0(t)$ according as t is or is not in $\boldsymbol{T}_1$. Then

$$(3.8) \qquad C_1(t) = \text{restriction of } C(t)E(t) \text{ to } \boldsymbol{H}_{1t} \text{ for } t \varepsilon \boldsymbol{T}_1,$$

because $I_1(t)$ = restriction of $E(t)$ to $\boldsymbol{H}_{1t}$ for $t \varepsilon \boldsymbol{T}_1$. Since $\boldsymbol{T}_1$ is a measurable set and $c(t)$ a measurable function, the operator-valued function $C(t)$ depends measurably on t. Hence there exists an operator C in $\boldsymbol{H}$ which decomposes into $C(t)$ under (3.1). Since $C(t)$ is a scalar for every t, and since (3.1) belongs by hypothesis to $\boldsymbol{Z}$ (which means that $\boldsymbol{Z}$ is exactly the ring of those operators in $\boldsymbol{H}$ which decompose into scalars under (3.11)), we have $C \varepsilon \boldsymbol{Z}$. But $Z \subseteqq \boldsymbol{W}$, and hence $C \varepsilon \boldsymbol{W}$. Also (3.8) implies that C_1 is the restriction of CE to $\boldsymbol{H}_1$. This proves $C_1 \varepsilon \boldsymbol{W}_1$, i.e. $\boldsymbol{P}_1 \subseteqq \boldsymbol{W}_1$. So we have proved $\boldsymbol{W}_1 \approx \sum \boldsymbol{W}_1(t)$ in the terminology of [15]. As remarked above, this fact proves that (3.3) is the central decomposition of $\boldsymbol{W}_1$. This proves the first assertion of Lemma 3.1.

But the fact (which we have just proved) that (3.3) is the central decomposition of $\boldsymbol{H}_1$ under $\boldsymbol{W}_1$ implies that $\boldsymbol{P}_1$ is the center of $\boldsymbol{W}_1$. On the other hand we have just seen that (3.8) implies that the elements C_1 of $\boldsymbol{P}_1$ are exactly the elements of the form "restriction of CE to $\boldsymbol{H}_1$" where C varies over $\boldsymbol{Z}$, i.e. the elements of $\boldsymbol{Z}_1$. This proves $\boldsymbol{Z}_1 = \boldsymbol{P}_1$; hence $\boldsymbol{Z}_1$ is the center of $\boldsymbol{W}_1$. This proves the second assertion of Lemma 3.1 and hence the proof of Lemma 3.1 is complete.

Later on we shall also require

LEMMA 3.2. *Let again $E = E^* = E^2$ and $E \varepsilon \boldsymbol{W}'$. Let C_E be the smallest projection $\varepsilon \boldsymbol{Z} = \boldsymbol{W} \cap \boldsymbol{W}'$ which satisfies $C_E \supseteqq E$. Then the restriction of the homomorphism $X \to XE$ to the subalgebra $\boldsymbol{W}C_E$ of $\boldsymbol{W}$ is an isomorphism of $\boldsymbol{W}C_E$ onto $\boldsymbol{W}E$. The kernel of the homomorphism $X \to XE$ is $\boldsymbol{W}(I - C_E)$.*

Proof. By hypothesis $C_E \supseteqq E$, i.e. $C_EE = E$; hence $\boldsymbol{W}E = (\boldsymbol{W}C_E)E$, which proves that the onto-assertion of the lemma is trivial. In order to prove that $\boldsymbol{W}C_E$ is mapped isomorphically under the homomorphism

$$X \to XE, \tag{3.9}$$

it is clearly sufficient to prove that $\boldsymbol{W}(I - C_E)$ is the kernel of (3.9). Let $C_E(t)$ be the operator valued function into which C_E decomposes under (3.1). Since $C_E \varepsilon \boldsymbol{Z}$, and since $\boldsymbol{Z}$ belongs to (3.1), we have as in the proof of Lemma 3.1 $C_E(t) = c(t)I(t)$, where $c(t)$ is a numerical essentially bounded measurable function defined for all t. Since C_E is a projection, we must have $c(t) = 1$ or 0 for almost all t. But $C_EE = E$ implies $C_E(t)E(t) = E(t)$ for almost all t. Since $E(t) \neq 0(t)$ for $t \varepsilon \boldsymbol{T}_1$, we must have $c(t) = 1$ for almost all $t \varepsilon T_1$. Since $E(t) = 0(t)$ for $t \notin T_1$, we must have $c(t) = 0$ for almost all $t \notin \boldsymbol{T}_1$, for otherwise C_E would not be minimal among the central projections $\supseteqq E$. This proves

$$C_B(t) = \begin{cases} I(t) & \text{for almost all } t \in \boldsymbol{T}_1 \\ 0(t) & \text{for almost all } t \notin \boldsymbol{T}_1. \end{cases} \tag{3.10}$$

Now suppose X is in the kernel of the homomorphism (3. 9), i. e. suppose $XE = 0$. This means $X(t)E(t) = 0(t)$ for almost all t under the decomposition (3. 1). If $t \in \boldsymbol{T}_1$, then Lemma 2. 2 implies for almost all t that $X(t)E(t) = 0(t)$ if and only if $X(t) = 0(t)$. If on the other hand $t \notin \boldsymbol{T}_1$, then $X(t)E(t) = 0(t)$ for *any* $X \in \boldsymbol{W}$. This proves that $XE = 0$ if and only if $X(t) = 0(t)$ for almost all $t \notin \boldsymbol{T}_1$. But by (3. 10) this last statement is the same as $XC_B = 0$, i. e. $X(I - C_B) = I$, which proves that the kernel of the homomorphism (3. 9) is exactly $\boldsymbol{W}(I - C_B)$.

In the course of the proof of Lemma 3. 2 we obtained the following

COROLLARY 3. 1. *Let C_B be the smallest projection $\in \boldsymbol{Z}$ which satisfies $C_B \supseteqq E$. Let $C_B(t)$ be the operator valued function into which C_B decomposes under the central decomposition* (3. 1). *Then $C_B(t)$ is given for almost all t by* (3. 10), *where $\boldsymbol{T}_1$ is again the set of these t for which $E(t) \neq 0(t)$.*

4. The central decomposition of the regular representation of *G*. In this section we consider the central decomposition of the regular representation of our separable locally compact unimodular group G. Denote by $\boldsymbol{R}$ the weakly closed self-adjoint operator algebra generated by the right translations $R(g)$ in $\mathfrak{L}_2(G)$, and by $\boldsymbol{L}$ the algebra generated by the left translations $L(g)$. Godement and Segal have shown that $\boldsymbol{L}$ and $\boldsymbol{R}$ are each other's commuting algebras:

$$\boldsymbol{L} = \boldsymbol{R}' \quad \text{and} \quad \boldsymbol{R} = \boldsymbol{L}'. \tag{4.1}$$

Denote by $\boldsymbol{Z}$ the center of $\boldsymbol{R}$: $\boldsymbol{Z} = \boldsymbol{R} \cap \boldsymbol{R}' = \boldsymbol{R} \cap \boldsymbol{L} = \boldsymbol{L} \cap \boldsymbol{L}'$, and let

$$\mathfrak{L}_2(G) = \int_\oplus \mathfrak{H}_t \tag{4.2}$$

be the central decomposition of $\mathfrak{L}_2(G)$ under $\boldsymbol{R}$ (or $\boldsymbol{L}$).

If $a(g)$ is an arbitrary complex valued Haar-Lebesgue-integrable function on G, let $\boldsymbol{R}_a$ be the operator acting on $\mathfrak{L}_2(G)$ defined by

$$R_a = \int_G a(g)R(g)\,dg. \tag{4.3}$$

Clearly R_a is an element of $\boldsymbol{R}$. Denote by $\boldsymbol{R}^{(1)}$ the subset of $\boldsymbol{R}$ of the elements R_a obtained from integrable functions $a(g)$, and by $\boldsymbol{R}^{(1,2)}$ the subset of those $R_a \in \boldsymbol{R}^{(1)}$ for which $a(g)$ is also square integrable. Under the decomposition (4. 2) there corresponds to the operator R_a an operator-valued function, say

$A(t)$, and to the element $a(g)$ of $\mathfrak{L}_2(G)$ a vector valued function $a(t)$, whenever $a(g) \varepsilon \mathfrak{L}_1(G) \cap \mathfrak{L}_2(G)$. Again $A(t)$ and $a(t)$ can be changed arbitrarily on sets of measure zero. However we have

LEMMA 4.1. *Let $a(g) \varepsilon \mathfrak{L}_1(G) \cap \mathfrak{L}_2(G)$. Then there exists one choice which assigns to every operator R_a a unique operator valued function $A(t)$ and to the "vector" $a(g)$ a unique vector valued function $a(t)$ such that for this choice the mapping*

$$A(t) \to a(t) \tag{4.4}$$

is for almost every t a one-one linear mapping of a certain weakly dense linear subspace $\boldsymbol{P}(t)$ of the operator algebra $\boldsymbol{R}(t)$ defined below, onto a dense linear subspace of $\mathfrak{H}_t$.

Proof. It is clear that $\boldsymbol{R}$ is generated by $\boldsymbol{R}^{(1,2)}$. It follows from p. 386 of [16] that there exists a sequence of functions $a_j(g) \varepsilon \mathfrak{L}_1(G) \cap \mathfrak{L}_2(G)$ $(j = 1, 2, \cdots)$ which are dense in $\mathfrak{L}_2(G)$ and for which the corresponding operators

$$R_{a_1}, R_{a_2}, \cdots \tag{4.4a}$$

generate the ring $\boldsymbol{R}$ as the smallest weakly closed self-adjoint operator algebra containing the operators (4.4a). Now choose for each operator in the sequence (4.4a) *one* operator valued function $A_j(t)$ into which it decomposes under our given direct integral. Then it follows from the proof of Theorem VI (on p. 459) of [15] that after one possible change on a set of measure zero the factors $\boldsymbol{R}(t)$ into which $\boldsymbol{R}$ decomposes may be assumed to be generated by the operators $A_1(t), A_2(t), \cdots$. Now let A be a (not necessarily commutative) polynomial

$$p(R_{a_{j_1}}, R_{a_{j_2}}, \cdots) \tag{4.4b}$$

in a finite number of the operators (4.4a), and define $A(t)$ to be equal to

$$p(A_{j_1}(t), A_{j_2}(t), \cdots). \tag{4.4c}$$

Denote the family of operators so obtained for each t by $\boldsymbol{P}(t)$. Then $\boldsymbol{P}(t)$ is a dense linear subspace of $\boldsymbol{R}(t)$ in the weak topology for operators.

Now consider all monomials in some of the operators (4.4a). Since to the product $R_a R_{a'}$ corresponds the convolution of the functions $a(g)$ and $a'(g)$, i. e. $R_a R_a = R_{a''}$ where $a''(g) = \int a(g\gamma^{-1}) a'(\gamma) d\gamma$ and where $a'' \varepsilon \mathfrak{L}_1(G)$ whenever a and $a' \varepsilon \mathfrak{L}_1(G)$ it follows (for instance from Lemma 7.1 of [8]) that $a \varepsilon \mathfrak{L}_1(G) \cap \mathfrak{L}_2(G)$ and $a' \varepsilon \mathfrak{L}_1(G) \cap \mathfrak{L}_2(G)$ imply $a'' \varepsilon \mathfrak{L}_1(G) \cap \mathfrak{L}_2(G)$. Hence there exists a sequence of functions $b_1(g), b_2(g), \cdots$ each $\varepsilon \mathfrak{L}_1(G)$

$\cap \mathfrak{L}_2(G)$ such that each of the above monomials is equal to one of the operators R_{b_m}.

Clearly the functions $b_m(g)$ are dense in $\mathfrak{L}_2(G)$ since the $a_j(g)$ are contained among the $b_m(g)$. Now choose for each $b_m(g)$ one vector valued function $b_m(t)$ into which it decomposes under (4.2). It is proved in § 6 of [15] that the finite linear combinations of the $b_m(t)$ form a dense linear subspace of $\mathfrak{H}_t$ for almost every t and hence, after one change on a set of measure zero, for all t. Now let $x(g)$ be an arbitrary finite linear combination of the $b_m(g)$ with complex coefficients c_{m_i}:

$$x(g) = \sum_{i=1}^{r} c_{m_i} b_{m_i}(g) \qquad (r \text{ arbitrary } < \infty). \tag{4.5}$$

Define the vector valued function $x(t)$ by

$$x(t) = \sum_{i=1}^{r} c_{m_i} b_{m_i}(t) \tag{4.6}$$

then $x(g)$ decomposes into $x(t)$ under (4.2). As remarked above the $x(t)$ form for each fixed t a dense linear subspace of $\mathfrak{H}_t$, and the operators $X(t)$ defined by

$$X(t) = \sum_{i=1}^{r} c_{m_i} B_{m_i}(t) \tag{4.7}$$

are exactly the polynomials $p(A_{j_1}(t), A_{j_2}(t), \cdots)$ defined above. Hence the $X(t)$ form the dense subalgebra $\boldsymbol{P}(t)$ of $\boldsymbol{R}(t)$. Therefore the proof of Lemma 4.1 will be complete if we prove the following:

The correspondence $X(t) \leftrightarrow x(t)$ between the elements (4.7) *of $\boldsymbol{P}(t)$ and the elements* (4.6) *of $\mathfrak{H}_t$ is a one-one linear mapping for almost every t.* Denote by T_t the (essentially unique) relative trace which exists for certain elements of the factor $\boldsymbol{R}(t)$ in accordance with [13]. According to Lemma 7.4 of [8] the factor $\boldsymbol{R}(t)$ will be of type I or II for every t (after a possible change on one set of measure zero which we assume to have been made). Then there exists a certain function $\alpha(t) > 0$ and $\leq \infty$ such that

$$\int x(g)\bar{y}(g)\,dg = \int \alpha(t) T_t(X(t)Y(t)^*)\,ds(t), \tag{4.8}$$

where $ds(t)$ refers to the measure used to define the direct integral (4.2). On the other hand $\int x(g)\bar{y}(g)\,dg = \int (x(t), y(t))\,ds(t)$. Now let C be any element of $\boldsymbol{R}$, it decomposes into a certain operator valued function, say $C(t)$, under (4.2), and we obtain

$$\begin{aligned} \int (Cx)(g)y(g)\,dg &= \int (C(t)x(t), y(t))\,ds(t) \\ &= \int \alpha(t) T_t(C(t)X(t)Y(t)^*)\,ds(t). \end{aligned} \tag{4.9}$$

In accordance with Theorem IV and VII of [15] we obtain, as C ranges over the center $\boldsymbol{Z}$ of $\boldsymbol{R}$, for the functions $C(t)$ *all* essentially bounded *scalar valued* functions $c(t)I(t)$. Hence for $C \varepsilon \boldsymbol{Z}$ we obtain

$$\int (c(t)x(t), y(t))\,ds(t) = \int \alpha(t)c(t)T_t(X(t)Y(t)^*)\,ds(t).$$

In this equation $c(t)$ can be the characteristic function of an arbitrary s-measurable set, which proves

$$(x(t), y(t)) = \alpha(t)T_t(X(t)Y(t)^*) \tag{4.10}$$

for all t outside of some set of measure zero which may depend on $x(g)$ and $y(g)$. However if we put $b_m(g)$ for $x(g)$ and $b_n(g)$ for $y(g)$, we get from (4.10)

$$(b_m(t), b_n(t)) = \alpha(t)T_t(B_m(t)B_n(t)^*), \tag{4.11}$$

for all t outside of one set of measure zero which is obtained by taking the union of the countably many sets corresponding to each pair m, n of integers. Hence if $x(g)$ and $y(g)$ are finite linear combinations of the functions $b_m(g)$, then we can define $X(t)$ and $Y(t)$ in terms of the functions $B_m(t)$ by equation (4.7), and $x(t)$ and $y(t)$ by equation (4.6), and obtain for this particular choice of the functions $x(t), y(t), X(t), Y(t)$ the truth of (4.10) for all t outside of one sea of measure zero.

It is proved in Theorem VIII of [15] that the above function $\alpha(t)$ (denoted there by $a(\lambda)$) is positive for all t. If we can prove that $\alpha(t)$ is finite for almost all t then we may consider the right side of equation (4.10) above to be an inner product defined on $\boldsymbol{P}(t)$. The left side of (4.10) is the inner product of the space $\mathfrak{H}_t$. Hence $\alpha(t) < \infty$ would imply that the correspondence $X(t) \leftrightarrow x(t)$ preserves inner products, hence is one-one and clearly linear for almost every t (and therefore for all t, after a change on one set of measure zero). Therefore the proof of Lemma 4.1 will be complete if we prove

LEMMA 4.2. *The function $\alpha(t)$ which occurs in the generalized Peter-Weyl-Plancherel formula* (4.8) *is finite and positive for almost every t.*

Proof. As remarked above, $\alpha(t) > 0$ was proved by von Neumann ([15], Theorem VIII). To prove $\alpha(t) < \infty$ we observe that for $m = n$ we obtain from (4.11)

$$\| b_n(t) \|^2 = \alpha(t)T_t(B_n(t)B_n(t)^*).$$

Since $\| b_n(t) \|^2 < \infty$, the expression $T_t(B_n(t)B_n(t)^*)$ would have to be 0 for any t for which $\alpha(t) = \infty$. But then

$$B_n(t) = 0(t). \tag{4.12}$$

The set of those t for which (4.12) holds is known to be measurable (cf. [15], § 13); hence so is the set of those t for which (4.12) holds for all $n = 1, 2, \cdots$. On the other hand the $B_n(t)$ generate the ring $\boldsymbol{R}(t)$, and $\boldsymbol{R}(t)$ contains the identity operator $I(t)$ for almost every t. Hence (4.12) can be true only on a t-set of measure zero. Therefore $\alpha(t) < \infty$ for almost every t as required. This completes the proof of Lemma 4.2 and hence also the proof of Lemma 4.1.

In the next section we shall require a special case of the following

LEMMA 4.3. *Let β be a countably additive complex valued set function on G. For $y(g)$, a suitable complex valued function on G, put*

$$(4.12) \qquad (R_\beta y)(g) = \int_G y(g\gamma^{-1})d\beta(\gamma) \text{ and } (L_\beta y)(g) = \int_G y(\gamma^{-1}g)d\beta(\gamma).$$

Assume β is such that R_β and L_β are bounded linear operators on $\mathfrak{L}_2(G)$ defined for all $y(g) \varepsilon \mathfrak{L}_2(G)$ and that R_β and L_β are defined for all $y(g) \varepsilon \mathfrak{L}_1(G)$ and satisfy $R_\beta \mathfrak{L}_1(G) \subset \mathfrak{L}_1(G)$ and $L_\beta \mathfrak{L}_1(G) \subset \mathfrak{L}_1(G)$. Denote by $R_\beta(t)$, $L_\beta(t)$ the operator valued functions into which the operators R_β, L_β respectively decompose under the central decomposition (4.2).

ASSERTION. *The mapping (4.4) can be taken to be such that under it there corresponds for almost every fixed t to the operator $A(t)R_\beta(t)$ the element $L_\beta(t)a(t)$ of $\mathfrak{H}_t$, and to the operator $R_\beta(t)A(t)$ the element $R_\beta(t)a(t)$ of $\mathfrak{H}_t$.*

Proof. Observe first that the definition (4.12) of R_β, together with the assumption that R_β be a bounded operator, implies $R_\beta \varepsilon \boldsymbol{R}$; similarly $L_\beta \varepsilon \boldsymbol{L}$. Hence R_β and L_β are decomposable operators under (4.2). Therefore the above function $R_\beta(t)$ and $L_\beta(t)$ exist.

Notice next that the definition (4.12) implies that

$$(4.13) \qquad R_a R_\beta = R_{L_\beta a} \quad \text{and} \quad R_\beta R_a = R_{R_\beta a}.$$

Hence for all t outside of a set of measure zero (which may depend on the function $a(g)$ and the measure β) we obtain

$$(4.14) \qquad R_a(t) R_\beta(t) = R_{L_\beta a}(t), \text{ and } R_\beta(t) R_a(t) = R_{R_\beta a}(t),$$

where $R_a(t), \cdots$ are arbitrary operator valued functions into which the operators $R_a, \cdots$ decompose under (4.2) (arbitrary in the sense that they may be changed arbitrarily on sets of measure zero). Now put in equations (4.14) the function $b_n(g)$ instead of $a(g)$ and replace the "arbitrary" $R_a(t)$ by the function $A(t)$ introduced in the proof of Lemma 4.1. By taking the union of countably many sets of measure zero we then obtain from (4.14)

$$A(t)R_\beta(t) = R_{L_\beta a}(t), \text{ and } R_\beta(t)A(t) = R_{R_\beta a}(t),$$

for all t outside of *one* set of measure zero whenever $a(g)$ is a finite linear combination of the functions $b_n(g)$. By hypothesis the functions $(L_\beta a)(g)$ and $(R_\beta a)(g)$ are elements of $L_1(G) \cap L_2(G)$ whenever $a(g) \varepsilon L_1(G) \cap L_2(G)$. Since the measure β is fixed, we may assume that the functions $b_n(g)$ are chosen such that $L_\beta b_n = b_{n'}$ and $R_\beta b_n = b_{n''}$ are again functions of the sequence $b_1(g), b_2(g), \cdots$. Put $(L_\beta a)(g) = {}_\beta a(g)$ and $(R_\beta a)(g) = a_\beta(g)$, and let $A_\beta(t)$, ${}_\beta A(t)$ be the operator-valued function of Lemma 4.1 into which $R_{R_\beta a}$, $R_{L_\beta a}$ decompose respectively. Then we get all t outside of *one* set of measure zero

$$A(t)R_\beta(t) = {}_\beta A(t), \text{ and } R_\beta(t)A(t) = A_\beta(t). \tag{4.15}$$

On the other hand $(L_\beta a)(g)$ decomposes into the vector valued function $(L_\beta a)(t) = {}_\beta a(t)$ say, for almost all t, and $R_\beta a$ into $A_\beta(t)a(t) = a_\beta(t)$ for almost all t, where ${}_\beta a(t)$ and $a_\beta(t)$ are the vector valued functions uniquely determined above by $a(g)$ for which Lemma 4.1 is true.

Moreover, since we may assume $L_\beta b_n = b_{n'}$ and $R_\beta b_n = b_{n''}$ we see that the operators ${}_\beta A(t)$ and $A_\beta(t)$ are elements of $P(t)$; therefore the mapping (4.4) is defined for ${}_\beta A(t)$ and $A_\beta(t)$, and for all t outside of *one* set of measure zero we obtain

$${}_\beta A(t) \to {}_\beta a(t) \text{ and } A_\beta(t) \to a_\beta(t),$$

under the mapping (4.4). Combining this with (4.15) we obtain the truth of Lemma 4.3.

5. Completion of the proof of Theorem 1. Let us now consider the repersentation $U = \operatorname{ind} u$ of G defined in 1. If the representation $u\colon k \to u(k)$ of K is a discrete direct sum of representations u_ν, $u(k) = \sum_\oplus u_\nu(k)$, then it follows that U is the direct sum of representation U_ν of G, where $U_\nu = \operatorname{ind} u_\nu$:

$$\operatorname{ind} \{ \textstyle\sum_\oplus {}_\nu u_\nu \} = \textstyle\sum_\oplus {}_\nu \operatorname{ind} u_\nu. \tag{5.1}$$

Indeed if the representation space $\mathfrak{h}$ of u is a direct sum $\sum_\oplus \mathfrak{h}_\nu$ of invariant subspace $\mathfrak{h}_\nu$, then consider those functions $X(g) \varepsilon \mathfrak{H}$ for which $X(g)$ is, for each fixed $g \varepsilon G$, an element of $\mathfrak{h}_\nu$. They form a closed invariant linear subspace $\mathfrak{H}_\nu$ of $\mathfrak{H}$ under the operators $U(g)$, and if $U_\nu(g)$ denotes the restriction of $U(g)$ to $\mathfrak{H}$, then the definition of induced representation as given in **1** implies that U_ν is (up to unitary equivalence) equal to $\operatorname{ind} u_\nu$. This proves (5.1).

In particular let λ be the left-regular representation of K. Then

$\lambda = \sum_{\oplus} h_\nu u_\nu$, where u_ν varies over all irreducible representations of the compact group K (equivalent representations being identified); u_ν is not equivalent to $u_{\nu'}$ for $\nu \neq \nu'$, h_ν denotes the degree of u_ν and $h_\nu u_\nu$ denotes the repetition of the representation u_ν, h_ν times. Hence (5.1) gives

$$\operatorname*{ind}_{K\uparrow G} \lambda = \sum_{\oplus} h_\nu \operatorname{ind} u_\nu. \tag{5.2}$$

On the other hand the definition of induced repersentation implies that if λ is the regular representation of K (i. e. $\lambda(k)$ is left translation by the element k in the space $\mathfrak{L}_2(K)$), then $L = \operatorname{ind} \lambda$ is the (left-) regular representation of G. To see this, observe that the definition of induced representation implies that the repersentation space $\mathfrak{H}_\lambda$ for $\operatorname{ind} \lambda$ can be identified with the space of those elements $x(k,g)$ of $\mathfrak{L}_2(K) \times \mathfrak{L}_2(G)$ which satisfy

$$x(k, k'g) = x(kk'^{-1}, g). \tag{5.3}$$

Hence the square norm equals

$$\int_K \int_G | x(k,g) |^2 dg dk = \int_K \{ \int_G | x(1, k^{-1}g) |^2 dg \} dk$$
$$= \int_K \{ \int_G | x(1,g) |^2 dg \} dk = \int_G | x(1,g) |^2 dg,$$

where we take the Haar measure dk on K to be normalized so that $\int_K dk = 1$. This proves that if $x(k,g) \, \varepsilon \, \mathfrak{H}_\lambda$, then the mapping

$$x(k,g) \to x(1,g) = x'(g) \tag{5.4}$$

is a unitary mapping of $\mathfrak{H}_\lambda$ onto $\mathfrak{L}_2(G)$. Also the mapping (5.4) clearly commutes with left translations by elements of G, which proves that $\operatorname{ind}\lambda$ is (up to unitary equivalence) equal to the left regular reprseentation L of G. The reader who is worried about the fact that the functions $x(k,g)$ are, as elements of $\mathfrak{L}_2(K) \times \mathfrak{L}.(G)$, only defined up to sets of measure zero, may restrict himself at first to functions $x(k,g)$ which are continuous in k, and then extend the isometry (5.4) uniquely to a unitary mapping from $\mathfrak{H}_\lambda$ onto $\mathfrak{L}_2(G)$.

Thus (5.2) can be written as

$$L = \sum_{\oplus} h_\nu U_\nu, \text{ where } U_\nu = \operatorname{ind} u_\nu. \tag{5.5}$$

Let χ_ν be the character of u_ν, i. e. $\chi_\nu(k) =$ trace of $u_\nu(k)$. Then the operator

$$\Lambda_{\chi_\nu} = h_\nu \int_K \chi_\nu(k) \lambda(k) \, dk = h_\nu \int_K \chi_\nu(k) \rho(k) \, dk$$

is a projection; here $\rho(k)$ denotes right-translation by the element k of K in the space $\mathfrak{L}_2(K)$. The subspace $\Lambda_{\chi_\nu} \mathfrak{L}_2(K)$ of $\mathfrak{L}_2(K)$ is representation space for the representation $h_\nu u_\nu$ of K, as is well known.

Therefore the space of those elements $x(k,g)$ of $\mathfrak{L}_1(K)\times\mathfrak{L}_2(G)$ which satisfy (5.3) and $h_\nu\int\chi_\nu(k)x(k'k^{-1},g)\,dk = x(k',g)$ can be taken to be representation space for $\operatorname{ind} h_\nu u_\nu$ under right translations by the elements of G. Under the unitary mapping (5.4) this space is mapped onto the subspace of all those elements $x'(g)$ of $\mathfrak{L}_2(G)$ which satisfy

$$(5.6)\qquad h_\nu \int_K \chi_\nu(k)x'(gk^{-1})\,dk = x'(g),$$
$$\text{i. e., } (R_{\chi_\nu}x')(g) = x'(g), \text{ if } R_{\chi^\nu} = h_\nu \int_K \chi_\nu(k)R(k)\,dk.$$

This proves

LEMMA 5.1. *Let u be any continuous irreducible (unitary) representation of degree h of the compact subgroup K and χ the character of u. Then the subspace $R_\chi\mathfrak{L}_2(G)$ is a direct sum of h subspaces each of which transforms under left translation by elements of G equivalently to the operators of the representation $U = \operatorname{ind} u$ of G.*[3]

Indeed, since K is compact, any continuous irreducible unitary representation u of K is equivalent to one of the above u_ν, as we have observed above.

Let us now keep the irreducible representation u of K with character χ fixed. Perform as in **4** the central decomposition (4.2) of the space $\mathfrak{L}_2(G)$ under the operators $R(g)$, $g\in G$. The operator R_χ defined by the last of equations (5.6) is a projection which commutes clearly with $L(g)$ for every $g\in G$, hence also with every element of the ring $\boldsymbol{L}$ which the operators $L(g)$ generate. Hence we may apply Lemma 3.1 and conclude that if t is an element of a certain measurable set (of positive measure) $\boldsymbol{T}_1$, then the space $R_\chi(t)\mathfrak{H}_t$ may be identified with the space $\mathfrak{H}_{1t}$ obtained by performing the central decomposition of the space $R_\chi\mathfrak{L}_2(G)$ under the operators $R_\chi L(G)$ of the representation $hU = \operatorname{ind}(hu)$. Here $\boldsymbol{T}_1$ is the set of those t for which $R_\chi(t) \not\equiv 0(t)$.

Put

$$(5.7)\qquad L_\chi = h\int_K \chi(k)L(k)\,dk,$$

and denote by $L_\chi(t)$ the operator-valued function into which the operator L_χ decomposes under the central decomposition (4.2).

[3] Note that it follows from this together with Lemma 7.1 of [8] that factors of type III cannot occur (except on a set of measure zero) in the central decomposition of $\operatorname{ind}_{K\uparrow G} u$, whenever K is compact and G unimodular and separable.

LEMMA 5.2. *Denote by* $\boldsymbol{R}(t)$ *the factor into which the ring* $\boldsymbol{R}$ *(generated by the right translation* $R(g)$*) decomposes for almost every t under the central decomposition* (4.2). *Then the restriction of the mapping* $X(t) \rightarrow x(t)$ *defined by* (4.4) *to those elements* $X(t)$ *which satisfy* $X(t) = R_\chi(t)X(t)R_\chi(t)$ *is for almost every t a one-one linear mapping between a dense linear subspace of* $R_\chi(t)\boldsymbol{R}(t)R_\chi(t)$ *and a dense linear subspace of the space* $R_\chi(t)L_\chi(t)\mathfrak{H}_t$. *Hence in particular*

$$\dim R_\chi(t)\boldsymbol{R}(t)R_\chi(t) = \dim R_\chi(t)L_\chi(t)\mathfrak{H}_t \tag{5.8}$$

where dim *denotes the (ordinary) dimension of a vector space over the field of complex numbers.*

Proof. Replace the measure β of Lemma 4.3 by the measure $h_\chi(k)\,dk$, where dk refers—as throughout—to the Haar-measure on K. Then the hypotheses of Lemma 4.3 are readily seen to be verfied by the operators R_χ and L_χ. Hence if $\boldsymbol{P}(t)$ denotes again the dense subalgebra of $\boldsymbol{R}(t)$ introduced in Lemma 4.1, then Lemma 4.3 tells us that under the mapping (4.4) the elements of $\boldsymbol{P}(t)R_\chi(t)$ are mapped into $L_\chi(t)\mathfrak{H}_t$, and the elements of $R_\chi(t)\boldsymbol{P}(t)$ into $R_\chi(t)\mathfrak{H}_t$. Hence $R_\chi(t)\boldsymbol{P}(t)R_\chi(t)$ is mapped into the subspace $R_\chi(t)L_\chi(t)\mathfrak{H}_t$ of $\mathfrak{H}_t$ (for almost every t). Denote by $\mathfrak{H}_t^0$ the image of $\boldsymbol{P}(t)$ (in $\mathfrak{H}_t$) under the mapping (4.4). Then $\mathfrak{H}_t^0$ is a dense subspace of $\mathfrak{H}_t$ and the image of $R_\chi(t)\boldsymbol{P}(t)R_\chi(t)$ under the mapping (4.4) consists—as we have just seen—of $R_\chi(t)L_\chi(t)\mathfrak{H}_t^0$. Hence clearly $R_\chi(t)\boldsymbol{P}(t)R_\chi(t)$ is a dense subalgebra (in the weak topology for operators) of $R_\chi(t)\boldsymbol{R}(t)R_\chi(t)$, and $R_\chi(t)L_\chi(t)\mathfrak{H}_t^0$ a dense linear subspace of $R_\chi(t)L_\chi(t)\mathfrak{H}_t$, which proves Lemma 5.2.

Let us now consider the space $R_\chi(t)\mathfrak{H}_t$. It is an invariant subspace of $\mathfrak{H}_t$ under the elements of the algebra $\mathfrak{L}(t)$ as follows from Lemma 13 of [15]. By Theorem 1.1 of [8] there exists a continuous unitary representation V_t of G, V_t: $g \rightarrow V_t(g)$, where the operators act in the space $\mathfrak{H}_t$ and generate the ring $\mathfrak{L}(t)$ according to Lemma 1.2 of [8]. So we have

$$R_\chi(t)V_t(g) = V_t(g)R_\chi(t)$$

for all g and all t outside of one set of measure zero which is independent of g. Therefore the operators $R_\chi(t)V_t(g)$ form a representation of G in the space $R_\chi(t)\mathfrak{H}_t$ for almost every t.

We shall now prove that

$$L_\chi(t) = h\int_K \chi(k)V_t(k)\,dk \tag{5.9}$$

for almost every t. Let $y_n(g)$ be a sequence of elements of $\mathfrak{L}_1(G)$ such that for the corresponding operators L_{y_n} acting on $\mathfrak{L}_2(G)$ we have strong $\lim L_{y_n} = I$.

Denote by $L_{y_n}(t)$ the operator valued function into which L_{y_n} decomposes under the central decomposition (4.2). In accordance with p. 442 of [15] there exist a subsequence of the y_n for which the $L_{y_n}(t)$ converge strongly to the identity $I(t)$ for almost every t. Let us assume that this subsequence has been chosen and denote it again by y_n. Since $L_{y_n} = \int_G y_n(g)L(g)dg$ by definition, we have for almost every t

$$L_{y_n}(t) = \int_G y_n(g)U(g,t)dg.$$

On the other hand it has been shown in § 1 of [8] that

$$L_{y_n}(t) = \int_G y_n(g)V_t(g)dg$$

for almost every t. Now put $z_n(g) = h\int_K y_n(gk^{-1})\chi(k)dk$; then

$$\text{strong lim } L_{z_n} = \text{strong lim } L_{y_n}L_\chi = L_\chi,$$

since $L_{z_n} = L_{y_n}L_\chi$. Also $L_{z_n}(t) = L_{y_n}(t)L_\chi(t)$ for almost every t implies

$$\begin{aligned}\operatorname*{strong\,lim}_{n\to\infty} L_{z_n}(t) &= \operatorname*{strong\,lim}_{n\to\infty}\int_G z_n(g)U(g,t)dg\\ &= \operatorname*{strong\,lim}_{n\to\infty}\int_G z_n(g)V_t(g)dg = L_\chi(t) = \int_K h\chi(k)U(k,t)dk.\end{aligned}$$

But since the operators $V_t(g)$ form a representation of G, the definition of $z_n(g)$ implies

$$\int z_n(g)V_t(g)dg \to L_\chi(t)h\int\chi(k)V_t(k)dk.$$

Hence strong lim $L_{y_n}(t) = I(t)$ for almost every t implies

$$\text{strong lim} \int z_n(g)V_t(g) = L_\chi(t).$$

for almost every t. This proves equation (5.9) for almost every t.

From (5.9) we infer

$$(5.10)\qquad R_\chi(t)L_\chi(t) = L_\chi(t)R_\chi(t) = h\int_K \chi(k)R_\chi(t)V_t(k)dk.$$

This proves

LEMMA 5.3. *The subspace $L_\chi(t)R_\chi(t)\mathfrak{H}_t$ of $R_\chi(t)\mathfrak{H}_t$ is for almost every t the sum of all those subspaces of $R_\chi(t)\mathfrak{H}_t$ which transform equivalently to the repersentation u of K under the operators $R_\chi(t)V_t(k)$, for $k \varepsilon K$. Here $R_\chi(t)V_t(k)$ is considered as an operator of the space $R_\chi(t)\mathfrak{H}_t$.*

Next we have

LEMMA 5.4. *The subalgebra $R_\chi(t)\mathbf{R}(t)R_\chi(t)$, considered as an algebra of operators of the space $R_\chi(t)\mathfrak{H}_t$ is for almost every t exactly the set of those*

bounded operators of the space $R_\chi(t)\mathfrak{H}_t$ which commute with (the restriction of) $R_\chi(t)V_t(g)$ (to the space $R_\chi(t)\mathfrak{H}_t$) for every $g \in G$.

Proof. According to (4. 1) the ring $\boldsymbol{R}$ is the commuting algebra of $\boldsymbol{L}$ in the space $\mathfrak{L}_2(G)$. Hence if we apply Lemma 13 of [15] to our above central decomposition of $\mathfrak{L}_2(G)$, we may conclude that $\boldsymbol{R}(t)$ is the commuting algebra of $\boldsymbol{L}(t)$ for almost every t. By § 1 of [8] we know that after omission of one t-test of measure zero we have $V_t(g) \in \boldsymbol{L}(t)$ for all $g \in G$, and that the $V_t(g)$ generate $\boldsymbol{L}(t)$. But all this together with $R_\chi(t) \in \boldsymbol{R}(t)$ implies Lemma 5. 4.

Let us now combine Lemmas 5. 1, 5. 2, 5. 3 and 5. 4. By Lemma 5. 3 we know that $L_\chi(t)R_\chi(t)\mathfrak{H}_t$ is the sum of all those subspaces of $R_\chi(t)\mathfrak{H}_t$ which transform equivalently to the given repersentation $u(k)$ of K under the operators $V_t(k)$ for $k \in K$. Hence we have

$$\text{(5. 11)} \quad \dim\,[L_\chi(t)R_\chi(t)\mathfrak{H}_t] = h \cdot [\text{multiplicity of } u \text{ in } R_\chi(t)V_t(K)].$$

Hence Lemmas 5. 2 and 5. 4 imply

$$\text{(5. 12)} \quad \dim\{[R_\chi(t)V_t(g)]'\} = h \cdot [\text{multiplicity of } u \text{ in } R_\chi(t)V_t(K)],$$

where $[R_\chi(t)V_t(g)]'$ denotes the commuting algebra of the operators $R_\chi(t)V_t(g)$ in the space $R_\chi(t)\mathfrak{H}_t$.

Now let us consider the operators $U(g)$ of the induced representation $U = \operatorname{ind} u$ acting in the space $\mathfrak{H}$ as defined in **1**. The representation u of K being irreducible, we can apply Lemma 5. 1 and conclude that the space $R_\chi\mathfrak{L}_2(G)$ can be identified with Kronecker product $\mathfrak{h} \times \mathfrak{H}$ such that $R_\chi L(g)$ becomes identified with $I_h \times U(g)$, where I_h denotes the identity matrix in the h-dimensional space $\mathfrak{h}$. If $\mathfrak{H} = \int_\oplus \mathcal{H}_t$ is the central decomposition of $\mathfrak{H}$ under the operators $U(g)$ introduced in **1** (cf. equation (1. 8)), then von Neumann's result on the essential uniqueness of the central decomposition (cf. loc. cit.) implies that the spaces obtained from the central decomposition of $R_\chi\mathfrak{L}_2(G)$ can be identified with the spaces $\mathfrak{h} \times \mathfrak{H}_t$ in such a manner that in particular each (measurable) operator-valued function in the one decomposition goes over into the corresponding operator-valued function in the other decomposition (neglecting of course sets of measure zero again). On the other hand we know from Lemma 3. 1 that the spaces obtained from the central decomposition of $R_\chi\mathfrak{L}_2(G)$ may be identified for $t \in \boldsymbol{T}_1$ with the spaces $R_\chi(t)\mathfrak{H}_t$, where the $\mathfrak{H}_t$ are the component spaces of the central decomposition of $\mathfrak{L}_2(G)$ itself, so that corresponding operator-valued functions go again over

into each other. Hence we see that there is a one-one correspondence $t' \leftrightarrow t$ between almost all elements t' which occur in (1.8) and almost all those elements t which occur in (4.2) and are elements of $\boldsymbol{T}_1$. Here $\boldsymbol{T}_1$ is the set of those t for which $R_\chi(t) \neq 0(t)$. Moreover this one-one correspondence $t' \leftrightarrow t$ is such that for each corresponding pair t', t there exists a unitary operator $J(t', t)$ mapping $\mathfrak{H}_t$, onto $R_\chi(t)\mathfrak{H}_t$ in such a manner that

$$(5.13) \qquad R_\chi(t) V_t(g) = J(t', t)[I_h \times \mathcal{V}_{t'}(g)] J(t', t)^{-1}.$$

Here the operators $\mathcal{V}_{t'}(g)$ are the operators of the unitary representation of G acting in $\mathcal{H}_{t'}$ as introduced in **1** (compare equations (1.8) and (1.9)).

From (5.13) we infer at once

$$(5.14) \qquad \dim\{[R_\chi(t) V_t(g)]'\} = h^2 \cdot \dim\{[\mathcal{V}_{t'}(g)]'\}$$

for the dimensions of the commuting algebras of the operators $R_\chi(t) V_t(g)$ and $V_{t'}(g)$ respectively.

Also (5.13) obviously implies

$$(5.15) \qquad h^2 \cdot [\text{multiplicity of } u \text{ in } \mathcal{V}_{t'}(K)] = \text{multiplicity of } u \text{ in } R_\chi(t) V_t(K).$$

Hence combining (5.12) and (5.15) we get

$$(5.16) \qquad \dim\{[R_\chi(t) V_t(g)'\} = h^2 \cdot [\text{multiplicity of } u \text{ in } \mathcal{V}_{t'}(K)]$$

for each corresponding pair t', t. Hence (5.14) and (5.16) together imply for almost every t'

$$(5.17) \qquad \text{multiplicity of } u \text{ in } \mathcal{V}_{t'}(K) = \dim\{[\mathcal{V}_{t'}(g)]'\}.$$

Since equation (5.17) is exactly the assertion of Theorem 1, i. e. equations (5.17) and (1.10) are identical, *the proof of Theorem* 1 *is herewith completed.*

Remark. In the course of this proof we have obtained somewhat more information than Theorem 1 asserts. For instance if we combine Lemmas 5.2, 5.3 and 5.4 we see that there is a *natural* linear one-one mapping between a dense linear subspace of the sum of those subspaces of $R_\chi(t)\mathfrak{H}_t$ which transform equivalently to $u(k)$ under the subgroup K and a dense linear subspace of the commuting algebra of the operators $R_\chi(t) V_t(g)$ in the space $R_\chi(t)\mathfrak{H}_t$.

THE JOHNS HOPKINS UNIVERSITY.

BIBLIOGRAPHY.

[1] E. Cartan, "Sur la determination d'un system orthogonal complet dans un espace de Riemann symmetrique clos," *Rendiconti del Circolo Matematico di Palermo*, vol. 53 (1929), pp. 217-252.

[2] ———, "Les espaces Riemannienes symetriques," *Verhandlungen des Internationalen Mathematiker-Kongresses*, Zürich, 1932, vol. 1, pp. 152-161.

[3] ———, "Sur les domaines bornés homogènes de l'espace de *n* variables complexes," *Abhandlungen aus dem Mathematischen Seminar der Hansischen Universität*, vol. 11 (1936), pp. 111-162.

[4] I. M. Gelfand, "Spherical functions in symmetric Riemann spaces," *Doklady Adademiia Nauk SSSR*, vol. 70 (1950), pp. 5-8.

[5] G. W. Mackey, "Imprimitivity for representations of locally compact groups. I," *Proceedings of the National Academy of Sciences*, vol. 35 (1949), pp. 537-545.

[6] ———, "Imprimitivité pour les representations des groupes localement compact II," *Comptes Rendus*, vol. 230 (1950), pp. 808-809, III., vol. 230 (1950), pp. 908-909.

[7] F. I. Mautner, "Unitary representations of locally compact groups I," *Annals of Mathematics*, vol. 51 (1950), pp. 1-25.

[8] ———, "Unitary representations of locally compact groups II," *ibid.*, vol. 52 (1950), pp. 528-556.

[9] ———, "The structure of the regular representation of certain discrete groups," *Duke Mathematical Journal*, vol. 17 (1950), pp. 437-441.

[10] ———, "Infinite dimensional irreducible representations of certain groups," *Proceedings of the American Mathematical Society*, vol. 1 (1950), pp. 582-584.

[11] ———, "On the decomposition of unitary representations of Lie groups," *Proceedings of the American Mathematical Society*, vol. 2 (1951), pp. 490-496.

[11a] ———, "Fourier analysis and symmetric spaces," *Proceedings of the National Academy of Sciences*, vol. 37 (1951), pp. 529-533.

[12] F. J. Murray and J. von Neumann, "Rings of operators," *Annals of Mathematics*, vol. 37 (1936), pp. 116-229.

[13] J. von Neumann, "On rings of operators III," *ibid.*, vol. 41 (1940), pp. 94-161.

[14] F. J. Murray and J. von Neumann, "On rings of operators IV," *Annals of Mathematics*, vol. 44 (1943), pp. 716-808.

[15] J. von Neumann, "On rings of operators. Reduction Theory," *ibid.*, vol. 50 (1949), pp. 401-485.

[16] ———, "Zur Algebra der Funktionaloperationen und Theorie der Normalen Operatoren," *Mathematische Annalen*, vol. 102 (1929), pp. 370-427.

[17] C. E. Rickart, "Banach algebras with an adjoint operation," *Annals of Mathematics*, vol. 47 (1946), pp. 528-549.

[18] C. L. Siegel, "Symplectic Geometry," *American Journal of Mathematics*, vol. 65 (1943), pp. 1-86.

[19] H. Weyl, "Harmonics on homogeneous manifolds," *Annals of Mathematics*, vol. 35 (1934), pp. 486-499.

[20] A. Wintner, "On the location of continuous spectra," *American Journal of Mathematics*, vol. 70 (1948), pp. 22-30.

CLASS GROUP RELATIONS IN CYCLOTOMIC FIELDS.[1]

By Robert E. MacKenzie.

1. Introduction. Let F be the cyclotomic extension of the field R of rational numbers obtained by the adjunction of a primitive n-th root of unity ϵ, $n \geqq 3$.

Let N be the multiplicative group of residue classes of the rational integers prime to n modulo n. If q is an element of N, then q will be regarded as either a residue class modulo n or as a rational integer representing this residue class, whichever is appropriate to the context. For each $q \varepsilon N$, let σ_q be the uniquely determined automorphism of F over R for which $\sigma_q(\epsilon^q) = \epsilon$. By this definition N may be identified with the Galois group of F over R.

The principal result contained in this paper is the following theorem.

Theorem 4. *Let $\mathfrak{K}$ be any ideal class of F. For all rational integers s and t,*

$$\prod_{q \varepsilon N} \sigma_q(\mathfrak{K})^{[q(s+t)/n]-[qs/n]-[qt/n]}$$

is the principal class. ($[x]$ indicates the greatest integer $\leqq$ the rational number x.)

This theorem lists a set of relations among the elements of the class group of any cyclotomic extension of the rational field. In the present state of algebraic number theory it is a rather isolated fact. It would be of considerable interest to know whether it has a more intimate connection with class field theory.

Theorem 4 is a generalization of a known theorem[2] for the case where n is a prime. The key to the proof consists in regarding the n-th power residue symbol[3] $(\alpha/\mathfrak{p})$ as a function on the additive group of the residue class field of the prime $\mathfrak{p}$ and employing the relation between the convolution of the functions $(\alpha/\mathfrak{p})^s$ and $(\alpha/\mathfrak{p})^t$ and the product of their Fourier trans-

[1] Received November 12, 1951.

[2] D. Hilbert, "Die Theorie der algebraischen Zahlkörper," *Jahresbericht der Deutschen Mathematiker Vereinigung*, vol. 4 (1897), § 109.

[3] H. Hasse, "Bericht über neurere Untersuchungen und Probleme aus der Theorie der algebraischen Zahlkörper," *ibid.*, Ergänzungsband VI (1930), § 10.

forms. Theorems 1 and 2 concern the factorization of the values of the Fourier transform of $(\alpha/\mathfrak{p})^s$. Theorem 3 gives the factorization of the values of the convolution of $(\alpha/\mathfrak{p})^s$ and $(\alpha/\mathfrak{p})^t$.

2. The Fourier transforms. Let $\mathfrak{p}$ be a prime ideal of the ring of integers of F which does not divide n. Let Φ be the additive group of the residue class field of $\mathfrak{p}$ and Φ^* the multiplicative group of non-zero residue classes. If α is an element of Φ (or Φ^*), then α will be regarded either as a residue class or as an integer which represents this residue class, whichever suits the context. $N\mathfrak{p}$ is the customary notation for the order of Φ.

Let X be the character group [4] of Φ^*. For $\chi \in X$ the Fourier transform ϕ_χ of χ is the function $\phi_\chi(\lambda) = \Sigma_\alpha \chi(\alpha)\lambda(\alpha)$, $\alpha \in \Phi^*$, defined on the characters [4] λ of Φ. Among the elements of X occur the functions $(\alpha/\mathfrak{p})^s$, where $(\alpha/\mathfrak{p})$ is the n-th power residue symbol [3] in F and s is a rational integer. When it requires special consideration, the Fourier transform of $(\alpha/\mathfrak{p})^s$ will be denoted by ϕ_s. For convenience, $(0/\mathfrak{p})$ will be set equal to 0. Then $(\alpha/\mathfrak{p})^s$ becomes a function defined on Φ.

THEOREM 1. *If* $\lambda \neq 1$, *then* $|\prod_{\chi \in X} \phi_\chi(\lambda)|^2 = N\mathfrak{p}^{N\mathfrak{p}-2}$.

Proof. Order the elements of Φ^* and of X in a fixed manner and consider the determinants

$$\Delta = \det\{\lambda(\alpha\beta) \mid \alpha, \beta \in \Phi^*\}$$

and

$$D = \det\{\chi(\alpha) \mid \chi \in X, \alpha \in \Phi^*\},$$

each consisting of $N\mathfrak{p} - 1$ rows and columns. Summing on α gives

$$D\Delta = \det\{\overline{\chi(\beta)}\ \Sigma_\alpha \chi(\alpha\beta)\lambda(\alpha\beta) \mid \chi \in X, \beta \in \Phi^*\},$$

where $\overline{\chi(\beta)}$ is the complex conjugate of $\chi(\beta)$. The sum occurring in this determinant is $\phi_\chi(\lambda)$ and hence $D\Delta = \bar{D}\Pi_\chi\, \phi_\chi(\lambda)$, $\chi \in X$.

D may be shown to be different from zero by forming

$$D\bar{D} = \det\{\Sigma_\chi \chi(\alpha)\overline{\chi(\beta)} \mid \alpha, \beta \in \Phi^*\}.$$

The sum occurring in this determinant is zero unless $\alpha = \beta$, in which case it is $N\mathfrak{p} - 1$. Therefore $D\bar{D} = (N\mathfrak{p} - 1)^{N\mathfrak{p}-1}$. It now follows that

$$|\Pi_\chi \phi_\chi(\lambda)| = |\Delta|.$$

---

[4] The values of the characters are roots of unity.

The theorem requires the evaluation of $|\Delta|^2$. Since

$$|\Delta| = |\det\{\lambda(-\nu\beta) \mid \nu, \beta \in \Phi^*\}|$$

it follows by suming on β that

$$|\Delta|^2 = |\det\{\Sigma_\beta \lambda(\{\alpha - \nu\}\beta) \mid \alpha, \nu \in \Phi^*\}|.$$

The sum appearing in this determinant is -1 unless $\alpha = \nu$, in which case it is $N\mathfrak{p} - 1$. That is to say,

$$|\Delta|^2 = |\det\{-1 + \delta_{\nu\mu} N\mathfrak{p} \mid \nu, \mu = 1, 2, \cdots, N\mathfrak{p} - 1\}|,$$

where $\delta_{\nu\mu}$ is the Kronecker delta. This determinant may be evaluated by well-known methods and gives the desired result.

Further considerations will be restricted to the special Fourier transforms ϕ_s. The values of $\lambda(\alpha)$ are p-th roots of unity, where p is the rational prime which $\mathfrak{p}$ divides. The values of $\phi_s(\lambda)$ are therefore integers in the extension E obtained by the adjunction of a primitive p-th root of unity to F. The problem is the factorization of the integers $\phi_s(\lambda)$ in the field E. Theorem 1 shows that the divisors of $\phi_s(\lambda)$ are necessarily divisors of p.

It will now be assumed that $\mathfrak{p}$ is a prime of first degree relative to R. This condition prevails throughout the remainder of the paper. Let $\mathfrak{p}_q = \sigma_q(\mathfrak{p})$, $q \in N$. According to Hilbert's theory of the factorization of primes in normal extensions,[5] the primes $\mathfrak{p}_q$ and $\mathfrak{p}_{q'}$ are distinct when $q \neq q'$ in N and as q runs through the elements of N the primes $\mathfrak{p}_q$ run through the divisors of p in F. It is not difficult to see[6] that $\mathfrak{p}_q$ is completely ramified in E, that is, there is a prime $\mathfrak{P}_q$ in E such that $\mathfrak{p}_q = \mathfrak{P}_q^{p-1}$. As q runs through the elements of N, the primes $\mathfrak{P}_q$ run through the divisors of p in E.

According to Theorem 1, the factorization of $\phi_s(\lambda)$ in E may be written $\phi_s(\lambda) = \Pi_q \mathfrak{P}_q^{\nu(s\|q)}$. The values of the exponents $\nu(s\|q)$ are given by the following theorem.

THEOREM 2. *If $\lambda \neq 1$, then $\nu(s\|q) = (p-1)\{1 - \omega(qs)/n\}$ if $s \not\equiv 0 \pmod n$ and $\nu(s\|q) = 0$ if $s \equiv 0 \pmod n$, where $\omega(m)$ is the least non-negative integer congruent to m $(\bmod n)$.*

Proof. The elements of Φ are represented by the rational integers $0, 1, 2, \cdots, p-1$, hence

$$\phi_s(\lambda) = \sum_{\nu=1}^{p-1} (\nu/\mathfrak{p})^s \lambda(\nu) = \sum_{\nu=1}^{p-1} (\nu/\mathfrak{p})^s \lambda^\nu(1).$$

---

[5] D. Hilbert, *loc. cit.*, Chap. X.

[6] D. Hilbert, *loc. cit.*, Chap. XXI.

Since $\lambda \neq 1$, $\lambda(1)$ is a primitive p-th root of unity. Let $\pi = \lambda(1) - 1$. Then[6] $\pi = \Pi_q \mathfrak{P}_q$, $q \varepsilon N$. Applying σ_q to the congruence[8] $(\nu/\mathfrak{p})^q \equiv \nu^{(p-1)q/n}$ $(\operatorname{mod} \mathfrak{p})$ gives $(\nu/\mathfrak{p}) \equiv \nu^{(p-1)q/n}$ $(\operatorname{mod} \mathfrak{p}_q)$ and

$$\phi_s(\lambda) \equiv \sum_{\nu=1}^{p-1} \nu^{(p-1)q/n}(1+\pi)^\nu \qquad (\operatorname{mod} \mathfrak{P}_q^{p-1}).$$

Since $\nu^{p-1} \equiv 1 \ (\operatorname{mod} p)$, one may replace the exponent $(p-1)qs/n$ by $(p-1)\omega(qs)/n$.

Using the binomial theorem,

$$\phi_s(\lambda) \equiv \sum_{\mu=1}^{p-1} \pi^\mu \sum_{\nu=\mu}^{p-1} \binom{\nu}{\mu} \nu^{(p-1)\omega(qs)/n} + \sum_{\nu=1}^{p-1} \nu^{(p-1)\omega(qs)/n} \qquad (\operatorname{mod} \mathfrak{P}_q^{p-1}).$$

If s is divisible by n, then $\phi_s(\lambda) = -1$ and $\nu(s \parallel q) = 0$. It may be assumed that $s \not\equiv 0 \ (\operatorname{mod} n)$. In this case the exponent $(p-1)\omega(qs)/\, n$ is not divisible by $p-1$, which means that the second sum in the congruence for $\phi_s(\lambda)$ is congruent to zero mod p. It may be dropped mod $\mathfrak{P}_q^{p-1}$.

Let $\mu \geqq 1$ and $1 \leqq l \leqq p-2$. It follows from $\binom{\nu}{\mu}\nu^l = (\nu^{\mu+l} +$ terms of lower degree in $\nu)/\mu!$ that

$$\sum_{\nu=\mu}^{p-1} \binom{\nu}{\mu} \nu^l \begin{cases} \equiv 0 \ (\operatorname{mod} p), \mu \leqq p-l-2 \\ \not\equiv 0 \ (\operatorname{mod} p), \mu = p-l-1. \end{cases}$$

Applying this to the congruence for $\phi_s(\lambda)$, one concludes that

$$\phi_s(\lambda) \equiv a\pi^{(p-1)\{1-\omega(qs)/n\}} \ (\operatorname{mod} \mathfrak{P}_q^{\,p-(p-1)\omega(qs)/n}),$$

where a is a rational integer $\not\equiv 0 \ (\operatorname{mod} p)$ and hence $\not\equiv 0 \ (\operatorname{mod} \mathfrak{P}_q)$. Theorem 2 now follows from this congruence.

3. The convolutions. The coefficient of $\lambda(\alpha)$ in the product $\phi_s(\lambda)\phi_t(\lambda)$ will be denoted by $g_s{}^t(\alpha)$. Its explicit value is $\Sigma_\beta(\beta/\mathfrak{p})^s(\{\alpha-\beta\}/\mathfrak{p})^t$, which is an integer in F. Owing to the multiplicative nature of the power residue symbol, $g_s{}^t$ is almost the function $(\alpha/\mathfrak{p})^{s+t}$.

Case 1. $\alpha \neq 0$.

$$g_s{}^t(\alpha) = \Sigma_\beta(\alpha\beta/\mathfrak{p})^s(\{\alpha-\alpha\beta\}/\mathfrak{p})^t = (\alpha/\mathfrak{p})^{s+t} g_s{}^t(1).$$

Case 2. $\alpha = 0$.

$$g_s{}^t(0) = (-1/\mathfrak{p})^t \Sigma_\beta(\beta/\mathfrak{p})^{s+t} = \begin{cases} 0 & , s+t \not\equiv 0 \ (\operatorname{mod} n) \\ (-1/\mathfrak{p})^t(N\mathfrak{p}-1), & s+t \equiv 0 \ (\operatorname{mod} n). \end{cases}$$

Using these values of $g_s{}^t(\alpha)$ it follows that

$$\phi_s(\lambda)\phi_t(\lambda) = \begin{cases} g_s{}^t(1)\phi_{s+t}(\lambda) & , s+t \not\equiv 0 \pmod n \\ g_s{}^t(1)\phi_{s+t}(\lambda) + (N\mathfrak{p}-1)(-1/\mathfrak{p})^t, & s+t \equiv 0 \pmod n. \end{cases}$$

The factorization of $g_s{}^t(1)$ in F may be written $g_s{}^t(1) = \Pi_q \mathfrak{p}_q{}^{\nu(s\mid t\parallel q)}$, as will be demonstrated in Theorem 3, which gives the values of the exponents $\nu(s \mid t \parallel q)$ in most cases.

THEOREM 3. *If neither s nor t is divisible by n, then*

$$\nu(s \mid t \parallel q) = 1 - [q(s+t)/n] + [qs/n] + [qt/n].$$

Proof. *Case* 1. $s + t \not\equiv 0 \pmod n$.

$$g_s{}^t(1) = \phi_s(\lambda)\phi_t(\lambda)/\phi_{s+t}(\lambda).$$

The results of Theorem 2 and the relations $\mathfrak{p}_q = \mathfrak{P}_q{}^{p-1}$ and

$$\omega(m) = m - [m/n]n$$

give the desired result.

Case 2. $s + t \equiv 0 \pmod n$. Put $\lambda = 1$ into the expression for $\phi_s(\lambda)\phi_t(\lambda)$. Then $g_s{}^t(1) = -(-1/\mathfrak{p})^t = \pm 1$ and $\nu(s \mid t \parallel q) = 0$.

4. The class group relations. The proof of Theorem 4 may now be given. Let $\mathfrak{K}$ be the ideal class which contains p. For this ideal class we may state from Theorem 3 that $\Pi_q\sigma_q(\mathfrak{K})^{\nu(s\mid t\parallel q)}$ is the principal class, provided that neither s nor t is divisible by n. Furthermore, the factorization $p = \Pi_q \mathfrak{p}_q$ shows that $\Pi_q\sigma_q(\mathfrak{K})$ is also the principal class. This means (1) that the first statement regarding $\mathfrak{K}$ is true even if s or t or both are divisible by n and (2) that

$$\Pi_q\sigma_q(\mathfrak{K})^{1-\nu(s\mid t\parallel q)}$$

is the principal class. The latter statement is Theorem 4 for the case when $\mathfrak{K}$ is the ideal class containing p.

The generalized theorem on arithmetic progressions [7] assures us that every ideal class of F contains not only one but an infinite number of primes satisfying the conditions on $\mathfrak{p}$ in this paper. Hence Theorem 4 is true for every ideal class of F.

INDIANA UNIVERSITY.

[7] For the most general statement of this theorem see H. Hasse, *ibid.*, vol. 35 (1926), § 8.

A THEOREM ON UNIFORM CAUCHY POINTS.*[1]

By W. L. Gordon and C. W. McArthur.

The purpose of this paper is to prove a theorem which implies the main results of recent papers of B. J. Pettis, [4], A. Alexiewicz [1], and A. M. Ostrowski [3] concerning the Osgood-Kuratowski-Banach theorem and some of its consequences. The terminology of [4] will be followed throughout this paper.

Let X be any topological space, and Φ any cardinal number. A subset E of X is a I_Φ set if it may be represented as the union of Φ nowhere dense sets; otherwise it is a II_Φ set; E is Φ-residual if its complement $X \setminus E$ is a I_Φ set. Let Y be any completely regular topological space, and of all uniformizations of the topology in Y [5], let $\{V_\alpha \mid \alpha \varepsilon A\}$ be one having least cardinal number $|A|$. For each element λ of a fixed directed set Λ, let f_λ be a function on X to Y. A point x_0 in X is a *Cauchy point* of $\{f_\lambda\}$ if (0.1) given any $\alpha \varepsilon A$, there is some λ_α such that $f_\lambda(x_0) \varepsilon V_\alpha(f_{\lambda_\alpha}(x_0))$ for all $\lambda \geqq \lambda_\alpha$. A point x_0 is a *uniform Cauchy point* of $\{f_\lambda\}$ if (0.2) for each $\alpha \varepsilon A$ there is some $\lambda_\alpha \varepsilon \Lambda$, and an open set G containing x_0, such that $f_\lambda(x) \varepsilon V_\alpha(f_{\lambda_\alpha}(x))$ for all $\lambda \geqq \lambda_\alpha$ and all $x \varepsilon G$. The set of Cauchy points of $\{f_\lambda\}$ will be denoted by C and the set of uniform Cauchy points by C_u. The closure and interior of any set E in a topological space will be denoted by E^* and E^0 respectively.

Our main result may now be stated.

Theorem 1. *Let $|\Lambda|$ represent the least of the cardinal numbers of cofinal subsets*[2] *of Λ. If each f_λ is continuous, and if $\Phi \geqq \max(|A|, |\Lambda|)$, then $C \setminus C_u$ is I_Φ.*

For each $\alpha \varepsilon A$ there exists ([5], p. 14) a non-negative real function $\sigma_\alpha(p, q)$ on $Y \times Y$ such that (i) $\sigma_\alpha(p, p) = 0$, (ii) $\sigma_\alpha(p, q) = \sigma_\alpha(q, p)$, (iii) if $\sigma_\alpha(p, q) < 1$ then $p \varepsilon V_\alpha(q)$ and $q \varepsilon V_\alpha(p)$, and (iv) given $\epsilon > 0$

* Received September 13, 1951.

[1] Part of this paper was written under contract N7-onr-434, Task Order III, Navy Department (the Office of Naval Research).

[2] Equivalently, let $|\Lambda|$ be the least of the cardinal numbers of domains of subnets of the net $\{f_\lambda\}$ in the sense of J. L. Kelley, "Convergence in topology," *Duke Mathematical Journal*, vol. 17 (1950), No. 3.

there is some $\beta \varepsilon A$ such that $|\sigma_\alpha(r,s) - \sigma_\alpha(p,q)| < \epsilon$ whenever $r \varepsilon V_\beta(p)$ and $s \varepsilon V_\beta(q)$. Let $M = [\mu]$ be a cofinal subset of Λ having cardinal number $|\Lambda|$, and for each $\alpha \varepsilon A$ and $\mu \varepsilon M$, let

$$Q_{\alpha\mu} = X[x: |\sigma_\alpha(f_\lambda(x), f_\mu(x))| \leq \tfrac{1}{2} \text{ for all } \lambda \geq \mu].$$

In [4] it is shown without using the continuity of the f_λ's that $C = \bigcap_\alpha \bigcup_\mu Q_{\alpha\mu}$, and $C_u = \bigcap_\alpha \bigcup_\mu Q^0{}_{\alpha\mu}$. From this it follows that

$$C \setminus C_u \subset E = \bigcup_\alpha \bigcup_\mu (Q_{\alpha\mu} \setminus Q^0{}_{\alpha\mu}).$$

By hypothesis, each f_λ is continuous; hence, by (iv) above, for fixed λ and μ, $|\sigma_\alpha(f_\lambda(x), f_\mu(x))|$ is a continuous function of x. Therefore each $Q_{\alpha\mu}$ is closed, and $Q_{\alpha\mu} \setminus Q^0{}_{\alpha\mu}$ is nowhere dense. Hence E is the union of $|A| \cdot |\Lambda|$ nowhere dense sets. If $|A| \cdot |\Lambda|$ is finite, E is nowhere dense and so obviously I_Φ. If $|A| \cdot |\Lambda|$ is infinite, then $|A| \cdot |\Lambda| = \max(|A|, |\Lambda|) \leq \Phi$, and E is necessarily a I_Φ set. As a subset of a I_Φ set, $C \setminus C_u$ is also I_Φ.

With the same hypotheses on the spaces X and Y as in definitions (0.1) and (0.2), and with $\{f_\lambda\}$ a directed family of functions on X to Y, a *point of almost equi-continuity* of $\{f_\lambda\}$ is a point x_0 satisfying the condition that (0.3), for each $\alpha \varepsilon A$ there exists a λ_α and an open set G containing x_0, such that $f_\lambda(x) \varepsilon V_\alpha(f_\lambda(x_0))$ for all $\lambda \geq \lambda_\alpha$ and all $x \varepsilon G$. The set of all such points will be denoted by C_a. We recall the following result ([4]).

LEMMA 1. *The relations $C_u \supset C^* \cap C_a = C \cap C_a$ always hold, and if f_λ is continuous for each λ in a cofinal subset of Λ, then $C_u = C^* \cap C_a = C \cap C_a$.*

The following corollary is a combined statement of Theorems 1 and 2 of Pettis ([4]), the latter part being an extension of a theorem of Alexiewicz ([1]).

COROLLARY 1.1. *If each f_λ is continuous, and $\Phi \geq \max(|A|, |\Lambda|)$, then $C_u \subset C_a$, and*

(1) *if C is Φ-residual, then C_u and C_a are Φ-residual;*

(2) *if C is a II_Φ set, then $C \cap C_a$ $(= C_u)$ is a II_Φ set.*

By Theorem 1, $C \setminus C_u$ is I_Φ, and since $X \setminus C_u = (C \setminus C_u) \cup (X \setminus C)$ is the union of two I_Φ sets, $X \setminus C_u$ is I_Φ; that is, C_u is Φ-residual. By Lemma 1, $C_u \subset C_a$, so C_a is also Φ-residual. The second part of the conclusion is evident since $C = (C \setminus C_u) \cup C_u$, and $C \setminus C_u$ is I_Φ.

Given any function f_0 on X to Y, let C_0 be the set of *points of continuity*

of f_0; and given f_0 and $\{f_\lambda\}$, let $C(f_0)$ and $C_u(f_0)$ represent the sets of those points for which (0.1) and (0.2) respectively hold when f_λ is replaced by f_0.

We recall the following lemma ([4]).

LEMMA 2. *For any f_0 and any $\{f_\lambda\}$, we have*

(1) $C \supset C(f_0)$; (2) $C_0 \supset C(f_0)^0 \cap C_a$;

(3) $C(f_0) \cap C_u \supset C_u(f_0) \supset C(f_0)^0 \cap C_u$;

(4) *if f_λ is continuous for each λ in a cofinal set, then $C_0 \supset C_u(f_0)$.*

The Osgood theorem now takes the following form.

THEOREM 2. *Given $\{f_\lambda\}$ and f_0, suppose that each f_λ is continuous and that $\Phi \geq \max\{|A|, |\Lambda|\}$. Then $C_u(f_0) \subset C_a \cap C_0$ and $C(f_0)^0 \setminus C_u(f_0)$ is I_Φ.*

From (1) and (3) of Lemma 2 we have $C_u(f_0) \subset C \cap C_u$, and from Lemma 1, $C \cap C_u = C \cap C_a$; thus $C_u(f_0) \subset C_a$. By (4) of Lemma 2, $C_u(f_0) \subset C_0$, hence we clearly have $C_u(f_0) \subset C_a \cap C_0$, or equivalently $C_u(f_0) = C_u(f_0) \cap C_a \cap C_0$. Using 3 of Lemma 2 we have

$$C(f_0)^0 \setminus C_u(f_0) \subset C(f_0)^0 \setminus (C(f_0)^0 \cap C_u) = C(f_0)^0 \setminus C_u \subset C \setminus C_u.$$

Hence by Theorem 1, $C(f_0)^0 \setminus C_u(f_0)$ is I_Φ.

COROLLARY 2.1. *Assuming the hypothesis of Theorem 2, then*

(i) *if $C(f_0)^0$ is Φ-residual then so is $C_u(f_0) \cap C_a \cap C_0 \cap C(f_0)^0$;*

(ii) *if $C(f_0)^0$ is II_Φ so is $C_u(f_0) \cap C_a \cap C_0 \cap C(f_0)^0$.*

From Theorem 2, $C_u(f_0) = C_u(f_0) \cap C_a \cap C_0$, and since

$$(X \setminus C(f_0)^0) \cup (C(f_0)^0 \setminus C_u(f_0)) = X \setminus (C(f_0)^0 \cap C_u(f_0)),$$

where $X \setminus C(f_0)^0$ and $C(f_0)^0 \setminus C_u(f_0)$ are I_Φ sets, (i) follows. To establish (ii) note that $C(f_0)^0 = (C(f_0)^0 \setminus C_u(f_0)) \cup (C_u(f_0) \cap C(f_0)^0)$, where $C(f_0)^0$ is II_Φ and $C(f_0)^0 \setminus C_u(f_0)$ is I_Φ. Therefore $C_u(f_0) \cap C(f_0)^0$ is II_Φ.

Observe that in the special case of Corollary 2.1 obtained by requiring $C(f_0) = X$, one may conclude in addition that $C_a = C_u = C_u(f_0) \subset C_0$, because in this case parts (3) and (4) of Lemma 2 imply $C_u = C_u(f_0) \subset C_0$ and Lemma 1 implies that $C_u = C_a$.

The set Λ is called *essentially denumerable* if $|\Lambda| \leq \aleph_0$ (e.g. positive integers). The following theorem summarizes the foregoing results in case Λ is essentially denumerable and Y is pseudo-metric. Parts (2), (3), and a consequence of (5) have been stated by Pettis [4] and in less generality by Alexiewicz [1].

THEOREM 3. *Let Y be pseudo-metric and Λ essentially denumerable, and suppose f_λ is continuous for each $\lambda \varepsilon \Lambda$. Then*

(1) *$C \setminus C_u$ is first category;*

(2) *if C is residual so are C_a and C_u;*

(3) *if C is second category so are C_a and C_u.*

If also f_0 is on X to Y, then

(4) *$C(f_0)^\circ \setminus C_u(f_0)$ is first category;*

(5) *if $C(f_0)^\circ$ is residual so is $C_u(f_0) \cap C_a \cap C_0 \cap C(f_0)^\circ$;*

(6) *if $C(f_0)^\circ$ is second category so is $C_u(f_0) \cap C_a \cap C_0 \cap C(f_0)^\circ$.*

If X is second category then so are C_u and C_a in cases (2), (3), and in the cases (5), (6) so is $C_u(f_0) \cap C_a \cap C_0 \cap C(f_0)^\circ$.

Following Bourbaki [2] we call X a *Baire Space* if every non-null open set of X is second category (e. g. a locally compact Hausdorff space or a complete pseudo-metric space). In such a case it is clear that any residual set is second category and everywhere dense. Hence

COROLLARY 3.1. *Let Y be a pseudo-metric space, X a Baire space, Λ essentially denumerable, and suppose f_λ is continuous for each $\lambda \varepsilon \Lambda$ and f_0 is on X to Y. If $C(f_0)^\circ$ is residual, then $C_u(f_0) \cap C_a \cap C_0 \cap C(f_0)^\circ$ is residual, second category, and everywhere dense.*

In a recent paper A. M. Ostrowski [3] states the following theorem: Let $f(t,x)$ be a continuous function of the point (t,x) for $a < x < b$ and $t \geq T$, and suppose we have $\lim_{t\to\infty} f(t,x) = f(x)$, where $f(x)$ is also continuous in (a,b); then for any $\epsilon > 0$ there exists a subinterval J of (a,b) and a number T_0 such that we have $| f(t,x) - f(x) | < \epsilon$ for $t \geq T_0$ and $x \varepsilon J$.

Using Corollary 3.1, Ostrowski's theorem may be strengthened to the following: Suppose for each $t \geq T$ that $f(t,x)$ is a continuous function on $X =$ real interval (a,b) into a pseudo-metric space Y, and that f is a function on X to Y such that $\lim_{t\to\infty} f(t,x) = f(x)$ for each $x \varepsilon X$. Then $C_a \cap C_u(f) \cap C_0$ is residual, second category, and everywhere dense, and $C_a = C_u(f) \subset C_0$.

TULANE UNIVERSITY.

REFERENCES.

[1] A. Alexiewicz, "On sequences of operations (I)," *Studia Mathematica*, vol. 11 (1949), pp. 1-30.

[2] N. Bourbaki, *Éléments de Mathématique*, Livre IV, Chapitre IX, Paris, 1948, p. 75.

[3] A. M. Ostrowski, "Generalization of a theorem of Osgood to the case of continuous approximation," *Proceedings of the American Mathematical Society*, vol. 5 (1950), pp. 648-649.

[4] B. J. Pettis, "Uniform Cauchy points and points of equicontinuity," *American Journal of Mathematics*, vol. 73 (1951), pp. 602-614.

[5] A. Weil, "Sur les espaces à structure uniforme et sur la topologie générale," *Actualités Scientifiques et Industrielles*, no. 551, Paris, 1937.

MATRICES WITH INTEGER ELEMENTS IN COMBINATORIAL INVESTIGATIONS.*

By H. J. Ryser.

1. **Introduction.** Let it be required to arrange n elements into n sets, such that each set contains exactly k distinct elements and such that every pair of sets has exactly $\lambda = k(k-1)/(n-1)$ elements in common $(0 < \lambda < k < n)$. The significance of this n, k, λ combinatorial problem has been described in detail in [2]. The problem for $n = N^2 + N + 1$, $k = N + 1$, $N \geqq 2$, and $\lambda = 1$ is equivalent to finding a projective plane with $N + 1$ points on a line. For $n = 4N - 1$, $k = 2N - 1$, and $\lambda = N - 1$, the problem reduces to a determination of an Hadamard matrix of order $4N$.

Let $x_1, \cdots, x_n$ denote the n elements and let $s_1, \cdots, s_n$ denote the n sets formed from these elements. The elements $x_1, \cdots, x_n$ may be listed in a row and the sets $s_1, \cdots, s_n$ in a column. One forms the incidence matrix A in the usual way by inserting 1 in row i and column j if the element x_j belongs to the set s_i and 0 otherwise. The matrix A of order n is composed entirely of zeros and ones and satisfies the matric equation $AA' = B$. Here A' denotes the transpose of A, and the martix B has k in the main diagonal and λ in all other positions. Moreover, for the incidence matrix A, one readily proves that $A'A = B$. Clearly, the matrix B is congruent to the identity, relative to the field of rational numbers. The utilization of this fact has resulted in a proof of the impossibility of the n, k, λ problem for large classes of n, k, and λ ([1], [2]).

Consider now two n by n symmetric matrices A and B, with integer elements. The matrix B is *integrally represented by A* provided there exists a matrix C with integer elements such that $C'AC = B$. In particular, B is integrally represented by the identity provided there exists an integral A such that $AA' = B$.

Let B now denote the n by n symmetric matrix with k in the main diagonal and $\lambda = k(k-1)/(n-1)$ in all other positions. If the n, k, λ problem has a solution, then clearly B must be integrally represented by the identity. The purpose of the present paper is to show that for certain n, k, and λ, the converse proposition is also valid.

* Received September 20, 1951.

Unfortunately, the classical problem of deciding whether or not one matrix is integrally represented by another is not entirely solved. An excellent account of this and related arithmetic topics may be found in [3].

2. Theorems concerning the n, k, λ problem. The following is a generalization of Theorem 3 derived in [1].

THEOREM 2.1. *Let B be the symmetric matrix of order n, with k in the main diagonal and $\lambda = k(k-1)/(n-1)$ in all other positions $(0 < \lambda < k < n)$. Suppose that $AA' = A'A = B$, where A has integral elements. Then A is composed entirely of zeros and ones, or entirely of zeros and minus ones. Consequently the matrix A yields a solution of the n, k, λ problem.*

Let s_i denote the sum of column i of the matrix $A = [a_{rs}]$. Then

$$\sum_{i=1}^{n} a_{ji} s_i = k^2, \tag{1}$$

where $j = 1, \cdots, n$. Adding, one obtains

$$\sum s_i^2 = nk^2. \tag{2}$$

It is clear from $A'A = B$ that $| s_j | = | \sum_i a_{ij} | \leqq \sum_i a_{ij}^2 = k$, whence $s_j^2 \leqq k^2$. But then (2) implies $| s_j | = k$. But if $s_j = k$, then all of the a's in the j-th column of A consist of zeros and ones. On the other hand, if $s_j = -k$, then all of the a's in the j-th column are zeros and minus ones. Thus the columns of A are either composed of zeros and ones, or zeros and minus ones. Since, however, the inner product of two column vectors of A is the positive integer λ, the matrix does not contain columns of both varieties.

Actually, the restriction of the normality of A in Theorem 2.1 is unnecessarily severe. However, any restriction imposed on the transforming matrix is undesirable. This difficulty is overcome subsequently for certain values of n, k, and λ.

THEOREM 2.2. *Let B denote the symmetric matrix of order n, with k in the main diagonal and $\lambda = k(k-1)/(n-1)$ in all other positions $(0 < \lambda < k < n)$. Suppose further that k and $k-\lambda$ are relatively prime, that $k-\lambda$ is odd, and that $AA' = B$, where A is integral. Then upon suitable multiplication of the columns of A by ± 1, the matrix A is transformed into an incidence matrix composed entirely of zeros and ones. Consequently A yields a solution of the n, k, λ problem.*

Again let s_i denote the sum of column i of A, and let t_{ij} denote the inner product of the i-th and j-th column vectors of A. We first derive a basic relationship connecting the quantities t_{ij} and the products $s_i s_j$, namely

$$k^2 t_{ij} = \lambda s_i s_j + k^2 (k - \lambda) \delta_{ij}, \tag{3}$$

where δ_{ij} is the Kronecker delta.

The matrix A is nonsingular, with determinant $| A | = \pm k(k-\lambda)^{(n-1)/2}$. For if column one of B is subtracted from each of the other columns, and if to row one there is then added each of the other rows, it readily follows that $| B | = k^2(k-\lambda)^{n-1}$. Thus $AA' = B$ implies that

$$A'A = A^{-1}BA. \tag{4}$$

Let A_{rs} denote the cofactor of the element a_{rs} of the matrix A. Then comparing elements in the (u, v) positions on both sides of (4), one obtains

$$| A | \, t_{uv} = \lambda \sum_{i,j} A_{iu} a_{jv} + (k-\lambda) \sum_i A_{iu} a_{iv}.$$

But $\sum_i A_{iu} a_{iv} = | A | \, \delta_{uv}$, and equation (1) implies that

$$| A | \, s_u = k^2 \sum_i A_{iu}. \tag{5}$$

Hence (3) follows.

Any column of A may be multiplied by -1 and the equation $AA' = B$ remains valid. Hence we may suppose without loss of generality that $s_i \geqq 0$. Now $| A | = \pm k(k-\lambda)^{(n-1)/2}$, and by hypothesis, k and $k-\lambda$ are relatively prime. Hence by equation (5), it follows that each $s_i \equiv 0 \bmod k$. Thus we may write $s_i = u_i k$. Moreover, from (3) it follows that

$$t_{ii} = \lambda u_i^2 + (k-\lambda). \tag{6}$$

Let us now suppose that some $u_i = 0$. Then $s_i = 0$, and by (6) it follows that $t_{ii} = k - \lambda$. But

$$0 \equiv s_i^2 \equiv a_{1i}^2 + \cdots + a_{ni}^2 \equiv k - \lambda (\bmod 2),$$

and this contradicts the assertion that $k-\lambda$ is odd. Hence each $u_i \neq 0$. But $s_1^2 + \cdots + s_n^2 = k^2 n$, whence $u_1^2 + \cdots + u_n^2 = n$. Hence it follows that each $u_i = 1$ and thus $s_i = k$ for all i. But then $t_{ii} = k$, and hence each a_{ij} must equal 0 or 1. Thus A is an incidence matrix.

The preceding theorem is of interest for the special case of projective planes and Hadamard matrices.

THEOREM 2.3. *Let B denote the symmetric matrix of order $n = N^2 + N + 1$, with $N + 1$ in the main diagonal and 1 in all other positions. Let N be an odd integer > 1. Then the assertion that there exists a projective plane with $N + 1$ points on a line is equivalent to the assertion that the martix B is integrally represented by the identity.*

THEOREM 2.4. *Let B denote the symmetric matrix of order $n = 4N - 1$, with $2N - 1$ in the main diagonal and $N - 1$ in all other positions. The assertion that Hadamard matrices of all orders congruent to zero modulo 4 exist is equivalent to the assertion that the matrix B is integrally represented by the identity, for every odd integer N.*

If B is integrally represented by the identity for every odd integer N, then by Theorem 2.2, there must exist Hadamard matrices of order $4N$. To construct the Hadamard matrix of arbitrary order congruent to zero modulo 4, it is only necessary to form the direct product with the Hadamard matrix whose order is a suitable power of 2 (*cf.* [4]).

Theorem 2.2 is not necessarily valid for $k - \lambda$ an even integer. In this case certain of the u_i may actually equal zero. If $u_1, u_2, \cdots, u_t$ constitute those u_i equal to zero, then equation (3) merely implies that $A'A$ is a direct sum of two matrices, one factor of which is $(k - \lambda)I$, where I is the identity matrix of order t.

To furnish an example, let H be an Hadamard matrix of order $N = 2$ or $N \equiv 0 \pmod 4$. The first column of H is to be composed entirely of plus ones. Consider now the matrix of order $N^2 + N$, which is the direct sum of the matrix H, taken $N + 1$ times. Let this matrix be bordered by a row vector

$$(1, 0, \cdots, 0, 1, 0, \cdots, 0, \cdots, 1, 0, \cdots, 0)$$

of $N^2 + N$ components, where each of the sequences of zeros contains $N - 1$ members. Let the resulting array be bordered by a column vector of $N^2 + N + 1$ components, consisting of 0 in the first position and 1 in all other positions. The resulting matrix A is of order $N^2 + N + 1$, and satisfies the matric equation $AA' = B$. However, the matrix A is not the incidence matrix of a finite projective plane.*

OHIO STATE UNIVERSITY.

* The writer expresses thanks to Dr. C. J. Everett, whose correspondence with the author provided considerable help and encouragement. Thanks are also due to the referee for improvements in the exposition.

BIBLIOGRAPHY.

[1] R. H. Bruck and H. J. Ryser, "The nonexistence of certain finite projective planes," *Canadian Journal of Mathematics*, vol. 1 (1949), pp. 88-93.

[2] S. Chowla and H. J. Ryser, "Combinatorial problems," *ibid.*, vol. 2 (1950), pp. 93-99.

[3] B. W. Jones, "The arithmetic theory of quadratic forms," *Carus Mathematical Monographs*, no. 10 (1950).

[4] R. E. A. C. Paley, "On orthogonal matrices," *Journal of Mathematics and Physics*, vol. 12 (1933), pp. 311-320.

A GENERAL THEORY OF RADICALS.*

I. Radicals in Complete Lattices.

By S. A. Amitsur.

The various radicals which have been hitherto defined by various authors (Artin, Levitzki, Jacobson, Brown-McCoy, etc.) constitute an important tool in the study of the structure of rings. The theory of radicals was recently extended to non associative and non distributive rings as well as to more general structures. The similarities which exist between some of these radicals (in the underlying definitions and reasoning) have been already observed by B. Brown and N. H. McCoy, and they developed in [4] a theory for radicals in groups which, in particular cases, yields some of the known radicals in the theory of rings, but bears no relations to others. The purpose of the present paper is to give an axiomatic study of radicals. In order to achieve the greatest possible generality it was found suitable to develop the theory of radicals for complete lattices.[1] The axiomatic approach and the general results obtained here will be applied in a subsequent paper where also the the results of [4] will be incorporated in our general theory.

One readily observes that each of the radicals which have been hitherto defined in rings is connected with some ring-property which is invariant under ring-homomorphism.[2] Thus the Jacobson-Perlis radical ([3]) grows up out of the property of quasi-regularity. Generally, by a π-radical N of a ring S is meant a maximal ideal N of S possessing a given property π of this type, and such that the quotient S/N is free of non zero ideals with the same property. It turns out that the theory of radicals is based on the following simple consequence of the homomorphism-invariance of the property π: If A, B, C, are any three ideals in a ring S such that the quotient ring A/B has the property π then $(A,C)/(B,C)$ has the same property.[3] This fact, and the method

* Received March 9, 1950; revised September 15, 1951.

[1] I am indebted to Prof. R. Baer for a remark which led to the present general treatment of the theory of radicals.

[2] Compare with [4], Theorem 2.

[3] This follows immediately since A/B is homomorphic with $A/A \cap (B,C) \cong (A.B,C)/(B,C) \cong (A,C)/(B,C)$.

which has been applied by Baer for defining his Lower Radical in [1] are the starting points of the theory of radicals developed in this paper.

It is found convenient to formulate our conditions and results not in terms of properties but rather in terms of binary relations ρ between elements of a complete lattice. Thus in the case of ideals in a ring S and a ring-property π we write $A \rho B$ if $A \supseteq B$ and the quotient ring A/B has the desired property π. In particular, a ρ-radical r of a lattice L is an element $r \varepsilon L$ such that $r \rho 0$ and $x \rho r$ implies $x = r$. In this notation the above mentioned condition may be formulated as follows: if $a \rho b$ then $(a \cup c) \rho (b \cup c)$.

In Section **1** the definition of Baer's lower radical is extended to complete lattices with binary relations of the preceding type. The discussion of the existence of the radical is carried out in Section **2**. In order to cover the problem of the connection between the radical of a ring and the radical of its ideals in the theory of radicals of associative rings the whole theory is extended in Section **4** to more general structures to be called complete pseudo-lattices. Some related questions and the dual development of the present theory is dealt with in Sections **5** and **6**. The last section deals with an application to lattices in which multiplication is defined. We obtain an extension of Baer's lower radical of [1] to such lattices. The dual definition of this radical yields the maximal idempotent element of such lattices. The latter is an extension of the idempotent kernel of rings and semi-groups defined by J. Levitzki in [2].

1. The upper radical.

Notations. Let M be a complete lattice. We denote by I_M and by 0_M the unit and the zero of M. When no confusions are expected, the subscript M will be omitted. Lattices and sublattices will always mean complete lattices and complete sublattices. By an M-interval $[a, b]$ is meant the set $\{x; x \varepsilon M, a \leq x \leq b\}$. The notation $\mathrm{Sup}[x; \cdots]$ and $\mathrm{Inf}[x; \cdots]$ will be used to denote the greatest lower bound and respectively the least upper bound of the elements x subjected to a condition which will replace the dots in the brackets. We refer to the relation $\geq$ of M as to an inclusion relation, and we say that a includes b if $a \geq b$.

We consider a set of sublattices $\{L\}$ of M such that if $a, b \varepsilon L$, where $L \varepsilon \{L\}$, the L-interval $[a, b]$ belongs also to the set $\{L\}$.

Definition 1. A binary relation ρ defined in M is called an *H-relation* in M if ρ satisfies:

(A) If $a \rho b$; $a, b \varepsilon M$, then $a \geq b$.

(B) $a \rho a$ for every $a \varepsilon M$.

(C) If $a \rho b$ and $c \geq b$ then $(a \cup c) \rho c$.

Remark. Let $a \rho b$. Since $b \cup c \geq b$, it follows by (C) that $(a \cup c) \rho ((b \cup c)$ for every $c \varepsilon M$.

When $a \rho b$ and $a, b \varepsilon L$ we write $a \rho b$ *in* L, and a is said to be a ρ-*element over* b *in* L. An element $a \varepsilon L$ is said to be a ρ-*element in* L if $a \rho 0_L$.

A sublattice L of M is said to be ρ-*semi simple* in M if L does not possess non zero ρ-elements, i. e. $x \rho 0_L$ in L holds only for $x = 0_L$. Let $a \leq b$ be two elements of L. If the L-interval $[a, b]$ is ρ-semi simple in M, we write $a \bar{\rho} b$ in L. The element b is said to be a $\bar{\rho}$-*element over* a *in* L. If $a \bar{\rho} I_L$, a is said to be a $\bar{\rho}$-*element in* L.

Definition 2. An element $r \varepsilon L$ is called a ρ-*radical* in L if r is both a ρ-element in L and a $\bar{\rho}$-element in L.

Example. Let M be the lattice of the ideals of an associative ring S. The relation ν in M defined to be: $a \nu b$ if $a^n \leq b \leq a$ for some integer n is readily seen to be an H-relation in M. In this example, the $\bar{\nu}$-elements are the radical-ideals defined by Baer in [1] and the ν-radical in M is the nilpotent radical of the ring S (in case it exists!).

Unless otherwise stated, binary relations to be considered hereafter will be H-relations and the sublattices of M will be restricted to the set $\{L\}$ considered above.

Define inductively the following chain of elements in L: $u_0(L, \rho) = 0_L$, $u_1(L, \rho) = \operatorname{Sup}[p; p \rho 0_L \text{ in } L]$, $u_\lambda(L, \rho) = \operatorname{Sup}[u_\nu(L, \rho); \nu < \lambda]$ for limit ordinal λ, and $u_\lambda(L, \rho) = \operatorname{Sup}[p; p \rho u_{\lambda-1}(L, \rho) \text{ in } L]$ for non limit ordinal λ. Thus the chain $\{u_\lambda\}$ [4] is a well-defined non-decreasing chain of elements of L.

LEMMA 1.1. *There is an ordinal τ such that $u_\tau(L, \rho) = u_\sigma(L, \rho)$ for every ordinal $\sigma \geq \tau$ and $u_\nu(L, \rho) < u_\mu(L, \rho)$ for $\nu < \mu \leq \tau$ (if $\tau > 1$).*

Since the chain $\{u_\lambda\}$ is a subset of L it is readily verified that there exists a minimal τ such that $u_\tau = u_{\tau+1}$. The rest of the lemma follows now immediately by the definition of the chain $\{u_\lambda\}$ and by the minimality of τ.

The element $u_\lambda(L, \rho)$ is called the λ-*th* ρ-*radical* of L, and u_τ, the element of the preceding lemma, is called the *upper* ρ-*radical* of L (in M). This element will be denoted by $u(L, \rho)$.[4]

[4] When no confusion is expected u_λ, will replace $u_\lambda(L, \rho)$. Similarly u and r will replace $u(L, \rho)$ and $r(L, \rho)$.

The object of the present section is to determine the relation between the upper ρ-radical of L and the $\bar{\rho}$-elements of L.

The following fact will be often used: if $a \bar{\rho} b$ in L, then $a \bar{\rho} b$ in every sublattice of L which contains both a and b.

The following lemma is fundamental:

LEMMA 1.2. *Let p, q be two elements of the L-interval $[a, b]$ such that $q \bar{\rho} b$ and $p \rho a$; then $q \geq p$.*

Proof. It follows by (C) that $(p \cup q) \rho q$ in L. Since $q \bar{\rho} b$ and $b \geq p \cup q \geq q$, $p \cup q = q$. That is $q \geq p$.

In particular if $a = I_L$ and $b = 0_L$, we have

COROLLARY 1.1. *Each $\bar{\rho}$-element in L includes every ρ-element of L.*

A simple consequence of the preceding corollary is:

THEOREM 1.1. *Every sublattice L possesses at most one ρ-radical.*

Indeed, if r_1 and r_2 are two ρ-radicals of L, then since $r_1 \bar{\rho} I$ and $r_2 \rho 0$, $r_1 \geq r_2$. Similarly $r_2 \geq r_1$. Thus $r_1 = r_2$.

If L possesses a ρ-radical, in the light of the preceding theorem we refer to this radical as *the ρ-radical* of L and denote it by $r(L, \rho)$.

Since $r(L, \rho)$ is a $\bar{\rho}$-element it follows by Corollary 1.1 that

COROLLARY 1.2. *If $r(L, \rho)$ exists in L, then $r(L, \rho)$ is the maximal ρ-element of L.*

LEMMA 1.3. *Let $Q = \{q\}$ be a set of $\bar{\rho}$-elements in L, then $t = \mathrm{Inf}[q; q \in Q]$ is also a $\bar{\rho}$-element in L.*

Let $p \in L$ be such that $p \rho t$. Since $q \geq t$, $q \in Q$, it follows by Lemma 1.2 that $q \geq p$. This holds for every $q \in Q$; hence $t \geq p$. This proves that t is a $\bar{\rho}$-element.

The set of all $\bar{\rho}$-elements of L is non vacuous, since evidently I_L is a $\bar{\rho}$-element in L. We obtain, therefore, by the preceding lemma,

COROLLARY 1.3. *The meet of all $\bar{\rho}$ elements of L is the minimal $\bar{\rho}$-element in L.*

The following is the main theorem of the present section.

THEOREM 1.2. *The upper ρ-radical $u(L, \rho)$ is the minimal $\bar{\rho}$-element of L.*

Proof. Since by Lemma 1.1 $u(L,\rho) = u_\tau = u_{\tau+1}$, it follows by definition of $u_{\tau+1}$ that $u(L,\rho)$ is a $\bar\rho$-element in L. Hence, if $m = \mathrm{Inf}[q; q$ is a $\bar\rho$-element in $L]$, $m \leq u(L,\rho)$. To prove that $u(L,\rho) \leq m$, it is sufficient to show in the light of Lemma 1.1 that $m \geq u_\nu(L,\rho)$ for every ordinal ν. By the preceding corollary m is a $\bar\rho$-element in L, hence by Corollary 1.1 $m \geq p$ for every ρ-element p in L. This yields $m \geq u_1(L,\rho)$. Let $m \geq u_\nu(L,\rho)$ for every ordinal $\nu < \lambda$. For limit ordinal λ it is evident that $m \geq u_\lambda$. If λ is not a limit ordinal, since $m \geq u_{\lambda-1}$ it follows by Lemma 1.2 that $m \geq p$ for every p which is a ρ-element over $u_{\lambda-1}$. This implies $m \geq u_\lambda$, and the proof is completed.

THEOREM 1.3. *A necessary and sufficient condition that $r(L,\rho)$ exist is that $u(L,\rho)$ is a ρ-element, and in this case $r(L,\rho) = u(L,\rho)$.*

Proof. In view of the preceding theorem, $u(L,\rho)$ is a $\bar\rho$-element. Hence, if $u(L,\rho)$ is a ρ-element, $u(L,\rho)$ is the ρ-radical of L. Conversely, let $r(L,\rho)$ exist in L. Since $r(L,\rho)\bar\rho I$ it follows by Theorem 1.2 that $r(L,\rho) \geq u(L,\rho)$. On the other hand, since $r(L,\rho)$ is a ρ-element, $u(L,\rho) \geq u_1(L,\rho) \geq r(L,\rho)$. Thus $r(L,\rho) = u(L,\rho)$, and the latter is therefore a ρ-element in L.

Since $r(L,\rho) = u_1(L,\rho)$ we have

COROLLARY 1.4. *If $r(L,\rho)$ exists, it is the maximal ρ-element of L.*

A relation between the upper ρ-radicals of two lattices is given in the following theorem:

THEOREM 1.4. *If $L \supseteq L'$ such that $0 = 0'$,*[5] *then $u(L,\rho) \geq u(L',p)$.*

The proof is achieved by showing inductively that $u(L,\rho) \geq u_\lambda(L',\rho)$. If $p\,\rho\,0$ in L', the same relation holds in L; hence $u(L,\rho) \geq u_1(L',\rho)$. For a limit ordinal λ it is evident that $u(L,\rho) \geq u_\lambda(L',\rho)$ if $u(L,\rho) \geq u_\nu(L',\rho)$ holds for every $\nu < \lambda$. Let λ be not a limit ordinal. If $p\,\rho\,u_{\lambda-1}(L',\rho)$, since p and $u_{\lambda-1}(L',\rho)$ belong to L, this relation holds also in L. Hence, from $u(L,\rho) \geq u_{\lambda-1}(L',\rho)$ it follows by Lemma 1.2 that $u(L,\rho) \geq p$. This yields that $u(L,\rho) \geq u_\lambda(L',\rho)$. q. e. d.

THEOREM 1.5. *Denote by L_a the L-interval $[a, I]$. If $u(L,\rho) \geq a$, then $u(L,\rho) = u(L_a,\rho)$.*

Since $u(L,\rho)\bar\rho I$ in L, the same holds in the interval L_a since $u(L,\rho) \geq a$.

[5] 0 and 0′ denote respectively the zeros of L and L'.

Hence, Theorem 1.2 yields $u(L,\rho) \geq u(L_a,\rho)$. The same reasoning implies that $u(L_a,\rho)\bar{\rho}I$ in L, since this holds in L_a; hence $u(L_a,\rho) \geq u(L,\rho)$ and the theorem is proved.

A relation between the upper radicals of two different relations ρ_1 and ρ_2 is discussed in

THEOREM 1.6. *If either* 1) $a\,\rho_1\,b$ *implies* $a\,\rho_2\,b$, *or* 2) $b\,\bar{\rho}_2\,I$ *in* L *implies* $b\,\bar{\rho}_1\,I$ *in* L, *holds then* $u(L,\rho_1) \leq u(L,\rho_2)$.

We first show that (1) implies (2). Indeed, if $x\,\bar{\rho}_2\,I$ holds and I is not a $\bar{\rho}_1$-element over x, then there is a $p \in L$ such that $p > x$ and $p\,\rho_1\,x$. Hence (1) implies $p\,\rho_2\,x$ which is a contradiction. Now assume that (2) holds. Since $u(L,\rho_2)\bar{\rho}_2 I$ it follows that $u(L,\rho_2)\bar{\rho}_1 I$; hence by Theorem 1.2 $u(L,\rho_1) \leq u(L,\rho_2)$.

Hence,

COROLLARY 1.5. *If* $a\,\rho_1\,b$ *implies* $a\,\rho_2\,b$, *and* $b\,\bar{\rho}_1\,I$ *implies* $b\,\bar{\rho}_2\,I$, *then* $u(L,\rho_1) = u(L,\rho_2)$.

2. The ρ-radical. This section deals with the existence of the ρ-radical $r(L,\rho)$. For $a \in L$, we denote by Q_a the L-interval $[0,a]$. If $u(Q_a,\rho) < a$, then $[u(Q_a,\rho),a]$ is a non zero ρ-semi simple L-interval. If $[x,a]$ is a non-zero ρ-semi simple L-interval in L, then $u(Q_a,\rho) \leq x < a$. Hence

LEMMA 2.1. $u(Q_a,\rho) = a$ *if and only if none of the non zero L-intervals* $[x,a]$, $x < a$, *is ρ-semi simple.*

Since $L \supseteq Q_a$ and both have the same zero, it follows by Theorem 1.4:

LEMMA 2.2. $u(L,\rho) \geq u(Q_a,\rho)$.

LEMMA 2.3. *If* $a\,\bar{\rho}\,b$ *in* L, *and* $b\,\bar{\rho}\,c$ *in* L, *then* $a\,\bar{\rho}\,c$ *in* L.

Indeed, if $c \geq p \geq a$ such that $p\,\rho\,a$, $(p \cup b)\rho\,b$. Since $c \geq p \cup b \geq b$ and $c\,\bar{\rho}\,b$, it follows that $p \cup b = b$, i.e., $b \geq p \geq a$. But then $a\,\bar{\rho}\,b$ implies $p = a$. This proves that $a\,\bar{\rho}\,c$ in L.

The main theorem of the present section is

THEOREM 2.1. *The ρ-radical exists in every lattice L of M if and only if ρ satisfies*

(D) *For every* $a, b \in L$ *such that* $a < b$ *and such that b is not a ρ-element over a, there is an element* $c \in L$, *such that* $c\,\bar{\rho}\,b$ *in* L.

Proof. Let ρ be an H-relation which satisfies (D). If $u(L,\rho)$ were not a ρ-element, then there is $u(L,\rho) > c \geq 0$ such that $c \bar{\rho} u(L,\rho)$. By Theorem 1.2 $u(L,\rho)\bar{\rho} I$ in L. Hence the preceding lemma yields $c \bar{\rho} I$. Thus $u(L,\rho) > c$ contradicts Theorem 1.2. This proves that $u(L,\rho)$ is a ρ-element, and therefore we obtain by Theorem 1.3 that $u(L,\rho) = r(L,\rho)$ and the latter exists.

Conversely, let ρ be an H-relation for which the ρ-radical exists in every lattice L (of the set $\{L\}$). Let $b > a$ and b be not a ρ-element over a in L. The L-interval $[a,b]$ belongs to the set $\{L\}$, hence its ρ-radical r exists. Evidently, r is the required element, i. e., $b > r \geq a$ and $r \bar{\rho} b$ in L.

Remark 2.1. From the first part of the preceding proof it follows that if (D) holds in some lattice L only for the element $a = 0$, then the ρ-radical $r(L,\rho)$ of this lattice L exists.

Definition 3. An H-relation which satisfies (D) is called an *R-relation.*

THEOREM 2.2. *Let ρ be an R-relation. Let $\{a_\alpha\}$ be a set of elements in L which are ρ-elements over b, $b \in L$; then* Sup a_α *is also a ρ-element over b.*

Proof. Let r be the ρ-radical of the L-interval $[b,a]$, where $a = \operatorname{Sup} a_\alpha$. Since $r \bar{\rho} a$, it follows by Lemma 1.2 that $r \geq a_\alpha$. Hence $r \geq a$. Thus $r = a$, and therefore $a \rho b$.

One may replace (D) by the conditions of the following theorem.

THEOREM 2.3. *The relation ρ is an R-relation if and only if ρ satisfies the following two conditions:*

(D_1) *If $a \rho b$ and $b \rho c$ then $a \rho c$ (Transitivity).*

(D_2) *If $a_1 \leq a_2 \leq \cdots$ is an ascending well ordered sequence of ρ-elements over b, then* Sup a_i *is also a ρ-element over b.*

Proof. Let ρ satisfy the conditions of the theorem. Applying (D_2) to the ρ-elements of the lattice L (in the case $b = 0_L$), one obtains readily by (D_2) the existence of a maximal ρ-element r. This element r is also a $\bar{\rho}$-element in L, for if $x \in L$, $x > r$ such that $x \rho r$ then since $r \rho 0$, it follows by (D_1) that $x \rho 0$ which contradicts the maximality of r. This proves that $r = r(L,\rho)$. Hence by Theorem 2.1 it follows that ρ satisfies (D) and, ρ is an R-relation.

Now let ρ be an R-relation. Condition (D_2) is a simple consequence of

Theorem 2. 2. To prove the validity of (D_1), let $r(C)$ be the ρ-radical of the M-interval $[c, a] = C$, where $a \rho b$ and $b \rho c$. Since $b \rho c$, $r(C) \geq b$, and we obtain, therefore, by Theorem 1. 5 that $r(C)$ is also the ρ-radical of the M-interval $[b, a]$. Since $a \rho b, r(C) \geq a$. Thus $r(C) = a$, which proves that $a \rho c$. This completes the proof of the theorem.

The reasoning of the first part of the preceding proof yield also the following corollaries.

COROLLARY 2. 1. *If ρ is any relation in M (not necessarily an H-relation) which satisfies* (A), (B), (D_1), *and condition* (D_2) *only in the special case* $b = 0_M$, *then M possesses maximal ρ-elements, and each maximal ρ-element is a ρ-radical in M.*

In particular, it follows by the preceding proof that if ρ is an H-relation which satisfies (D_1), then every sublattice L either possesses the ρ-radical or does not contain maximal ρ-elements. Hence

COROLLARY 2. 2. *If ρ is an H-relation which satisfies* (D_1), *then every lattice L which satisfies the ascending chain condition for ρ-elements possesses a ρ-radical.*

Example. Let M be the lattice of the ideals of an associative ring S. Consider the following relation between the iedals of S: $a \rho b$ if the quotient ring a/b is quasi regular in the sense of Jacobson ([3]). One readily proves that ρ is an H-relation which satisfies (D_1) and (D_2). Thus the theorems of the present section yield the existence of the Jacobson-Perlis radical.

It is readily seen that the H-relation ν defined in the preceding section satisfies (D_1) and need not satisfy (D_2).

Let L be a fixed sublattice of M. Suppose we distinguish among the elements of L a class Σ, which contains the unity I, of special elements of L. We prove that:

THEOREM 2. 4. *If for every $\bar{\rho}$-element a in L which is not a ρ-element, there exists a special $\bar{\rho}$-element q of Σ such that $a > a \cap q$ then the ρ-radical of L exists and it is the meet of all the special $\bar{\rho}$-elements of Σ.*

Proof. Let $m = \mathrm{Inf}[q; q \in \Sigma, q$ is a $\bar{\rho}$-element in $L]$. Since $I \in \Sigma$, m is well defined. It follows by Lemma 1. 3 that m is a $\bar{\rho}$-element in L. If m is not a ρ-element, then $m > m \cap q$ for some $\bar{\rho}$-element q of Σ. But this contradicts the definition of m. Hence m is the ρ-radical of L. q. e. d.

3. Complete pseudo-lattices. There are still some general aspects in the theory of radicals of rings whose generalization is not covered by the theory developed in the preceding sections. This section deals with one of these problems, the relation between the radical of a ring and the radicals of its ideals. To this end the notion of complete pseudo-lattices is introduced.

Definition 4. A set $\mathfrak{M}$ is called a complete *pseudo lattice* if a binary relation $\geq$ (inclusion) is defined in $\mathfrak{M}$ which satisfies

1) $0 \varepsilon \mathfrak{M}$ such that $x \geq 0$ for every $x \varepsilon \mathfrak{M}$.

2) Every $\mathfrak{M}$-interval $[a, b] = \{x; x \varepsilon \mathfrak{M}, a \leq x \leq b)\}$ is a complete lattice with regard to the inclusion relation.

Note that the relation $\geq$ defined in $\mathfrak{M}$ must be reflexive and anti-symmetric but need not be transitive, as can be seen by the following example: Let $\mathfrak{M}$ be the set of all subrings of a ring S. For $a, b \varepsilon \mathfrak{M}$, we write $a \geq b$ if b is an *ideal* in the subring a of S. The interval $[b, a]$ of $\mathfrak{M}$ can be identified with the complete lattice of the ideals of the quotient ring a/b.

By an *H-relation ρ in $\mathfrak{M}$* we mean a relation ρ defined in $\mathfrak{M}$ which is an H-relation in every $\mathfrak{M}$-interval and satisfies the condition

(C_1) If $a \cup b$, $a \cap b$ are defined in $\mathfrak{M}$, then $a \bar{\rho} (a \cup b)$ implies $(a \cap b) \bar{\rho}\, b$ in $\mathfrak{M}$.

For complete lattices $\mathfrak{M}$, condition (C_1) is a consequence of (C). Indeed if $(a \cap b) \bar{\rho}\, b$ does not hold in $\mathfrak{M}$, then $p\, \rho (a \cap b)$ for some $b \geq p > a \cap b$. Hence by (C) $(p \cup a) \rho\, a$, which implies $p \cup a = a$, i. e., $a \geq p$. This together with $b \geq p$ contradicts $p > a \cap b$. This proof does not work in the general case, since $p \cup a$ may not exist in $\mathfrak{M}$.

We denote by $r(n, m)$ the ρ-radical (if it exists) of the $\mathfrak{M}$-interval $[n, m]$. The aim of the present section is to prove

THEOREM 3. 1. *If $r(n, m)$ exists, then $r(q, m)$ exists for every $q \varepsilon [m, n]$, and $r(q, m) = q \cap r(n, m)$ if and only if ρ satisfies*

(E_1) *If $a \rho b$ then $c \rho b$ for every $c \varepsilon [b, a]$.*[6]

(E_2) *If $a \bar{\rho} b$ then $a \bar{\rho} c$ for every $c \varepsilon [a, b]$.*

Proof. Suppose ρ satisfies the requirements of this theorem. Let $r = r(n, m)$ exist. Since $r \geq q \cap r \geq n$ and $r \rho n$, (E_1) yields $(q \cap r) \rho\, n$.

[6] Note that if $c \varepsilon [b, a]$, the $\mathfrak{M}$-interval $[b, c]$ is not necessarily a subinterval of $[b, a]$.

Now $m \geq q \cup r \geq r$ and $r \bar{\rho} m$; hence (E_2) yields $r \bar{\rho}(q \cup r)$, which implies by (C_1) that $(q \cap r) \bar{\rho} q$. This proves that $q \cap r(n, m)$ is the ρ-radical $r(q, n)$.

The necessity of (E_1) and (E_2) follows immediately since both conditions are the particular cases $r(b, a) = a$ and $r(a, b) = a$ of the statement of the theorem.

4. The mapping $a \to r(a)$. Let L be a sublattice of M, and let ρ be an R-relation in M. For every $a \varepsilon L$ we denote by $r(a)$ the ρ-radical of the L-interval $[a, I]$. The correspondence $a \to r(a)$ is encountered in many problems in the theory of rings, e. g. the correspondence between primary ideals and their prime ideals. A general property of this correspondence is treated in this section.

THEOREM 4. 1. *Let ρ be an R-relation satisfying* (E_1) *and the condition* (F) $(a \cup b) \rho b$ *implies* $a \rho (a \cap b)$. *Then the correspondence $a \to r(a)$ is an idempotent meet homomorphism of L onto itself.*

Proof. The mapping $a \to r(a)$ is isotone. For let $a \geq b$. Since $r(a) \bar{\rho} I$ in the L-interval $[a, I]$, the same holds in the L-interval $[b, I]$. Thus Theorem 1. 2 yields $r(a) \geq r(b)$.

Now let $a_1, a_2 \varepsilon L$. Put $r_1 = r(a_1)$ and $r_2 = r(a_2)$. Since $a_1 \geq a_1 \cap a_2$, $r_1 \geq r(a_1 \cap a_2)$, and therefore $r_1 \cap r_2 \geq r(a_1 \cap a_2)$. Since $r_1 \geq (r_1 \cap r_2) \cup a_1 \geq a_1$, and since $r_1 \rho a_1$ it follows by (E_1) that $((r_1 \cap r_2) \cup a_1) \rho a_1$. The condition of the theorem implies $(r_1 \cap r_2) \rho ((r_1 \cap r_2) \cap a_1) = r_2 \cap a_1$. Similarly, since $r_2 \geq (a_1 \cap r_2) \cup a_2 \geq a_2$ and $r_2 \rho a_2$, $(a_1 \cap r_2) \cup a_2) \bar{\rho} a_2$, and hence $(a_1 \cap r_2) \rho ((a_1 \cap r_2) \cap a_2) = a_1 \cap a_2$. We obtain $(r_1 \cap r_2) \rho (a_1 \cap a_2)$ and $(a_1 \cap r_2) \rho (a_1 \cap a_2)$. Hence it follows by (D_1) that $(r_1 \cap r_2) \rho (a_1 \cap a_2)$. This yields $r(a_1 \cap a_2) \geq r_1 \cap r_2$. Hence $r(a_1) \cap r(a_2) = r(a_1 \cap a_2)$. Since $r(a) \bar{\rho} I$, we have $r(r(a)) = r(a)$, which proves the idempotency of this mapping.

5. Dual relations. On account of the duality of lattices, we develop in this section a theory of dual relations and dual radicals.

Definition 5. A relation σ defined in a lattice M is called a *dual R-relation* if σ satisfies:

(A′) If $a \sigma b$, $a, b \varepsilon M$ then $a \leq b$.

(B′) = (B) $a \sigma a$ for every $a \varepsilon M$.

(C′) If $a \sigma b$ and $c \leq b$ then $(a \cap c) \sigma c$.

Dually to the non decreasing chain of elements $\{u_\lambda\}$ we define a non increasing chain of radicals $\{l_\lambda\}$ as follows: $l_1(L,\sigma) = \operatorname{Inf}[q; q\,\sigma\, I \text{ in } L]$. $l_\lambda(L,\sigma) = \operatorname{Inf}[l_\nu; \nu < \lambda]$ for limit ordinal λ, and $l_\lambda(L,\sigma) = \operatorname{Inf}[q; q\,\sigma\, l_{\lambda-1}$ in $L]$ for non limit ordinal λ. The element $l_\lambda(L,\sigma)$ is called the λ-th *σ-radical of L.* This non increasing chain of radicals terminates at some $l_\tau(L,\sigma)$, that is, there is a minimal τ such that $l_\tau = l_\mu$ for every $\mu \geq \tau$. The element $l_\tau(L,\sigma) = L(L,\sigma)$ will be called the *lower σ-radical* of L.

One readily develops a theory for the lower radical dually to the theory of the upper radical developed in the preceding sections. In particular we refer to a dual H-relation as a *dual R-relation* if it satisfies

(D′) *If $a\,\sigma\, b$ does not hold, then $c\,\bar{\sigma}\, b$ does not hold, then $c\,\bar{\sigma}\, b$ for some $c \in L$, $a \leq c < b$.*

Let ρ be an H-relation defined in M. The relation $\bar{\rho}$ defined in M can be considered as a dual relation in M. We have

THEOREM 5.1. *If ρ is an H-relation in M, then $\bar{\rho}$ is a dual H-relation which satisfies* (D′$_1$).[7] *Furthermore, if the set L of the sublattices of M is the set of the M-intervals, then $\bar{\rho}$ is a dual R-relation, and $u(L,\rho) = r(L,\bar{\rho})$ for every M-interval L.*

Proof. Evidently $\bar{\rho}$ satisfies (A′) and (B′). We proceed in proving (C′). Let $a\,\bar{\rho}\, b$ and $b \geq c$. If $c\,\bar{\rho}(a \cap c)$ is not true, then $p\,\rho(a \cap c)$ for some $c \geq p > c \cap a$. Since ρ satisfies (C), $(a \cup p)\,\rho\, a$. But in view of $b \geq p \cup a \geq a$ and $a\,\bar{\rho}\, b$, it follows that $p \cup a = a$, i. e., $a \geq p$. This together with $c \geq p$ contradicts $p > c \cap a$, and thus (C′) is proved. Condition (D′$_1$) was proved in Lemma 2.3.

Before proceeding with the proof we remark that if $\{L\}$ is the totality of the M-intervals of M, then $a\,\bar{\rho}\, b$ in an M-interval L is the same requirement as $a\,\bar{\rho}\, b$ in the whole lattice M.

We prove now that the upper ρ-radical $u(L,\rho)$ of an M-interval L is the $\bar{\rho}$-radical of L. By Theorem 1.2 it follows that $u(L,\rho)\bar{\rho}\, I_L$ in L, and hence $u(L,\rho)\bar{\rho}\, I_L$ in M. The interval $[0_L, u(L,\rho)]$ is $\bar{\rho}$-semi simple, for otherwise $q\,\bar{\rho}\, u(L,\rho)$ for some $0_L \leq q < u(L,\rho)$, $q \in L$. Hence Lemma 2.3 implies $q\,\bar{\rho}\, I_L$. This contradicts the minimality of $u(L,\rho)$ proved in Theorem 1.2. This shows that $u(L,\rho) = r(L,\bar{\rho})$, which, by the dual of Theorem 2.1, proves that $\bar{\rho}$ is a dual R-relation.

---

[7] (D′$_1$) denotes the dual condition of (D$_1$).

In the light of the preceding theorem, Theorem 1.2 is the dual of Corollary 1.4.

Some of the known relations are at the same time both H-relations and dual H-relations. The following theorem deals with a class of such relations.

THEOREM 5.2. *If ρ is any relation defined in M which satisfies* (A), (B), (E_1) *and the requirement that* $(a \cup b)\rho\, a$ *implies* $b\, \rho(a \cap b)$, *then the relation* ρ' *defined by* $a\, \rho'\, b$ *if* $b\, \rho\, a$, *is a dual H-relation.*

Evidently ρ' satisfies (A′) and (B′). Let $a\, \rho'\, b$ and $b \geq c$. Since $b \geq c \cup a \geq a$ and $b\, \rho\, a$, it follows by (E_1) that $(c \cup a)\rho\, a$ and, therefore, $c\, \rho(a \cap c)$. This proves that $(a \cap c)\rho'\, c$, hence the validity of (C′).

6. Multiplicative lattices. We conclude this paper with an example which generalizes two known radicals, the lower radical defined by Baer in [1], and the idempotent kernel of rings defined by Levitzki in [2].

By a multiplicative lattice M we mean a complete lattice in which multiplication is defined satisfying

1) For every $a, b \in M$, $ab \in M$ is uniquely defined,

2) $ab \subseteq a \cap b$,

3) $a(b \cup c) = ab \cup ac$; $\quad (b \cup c)a = ba \cup ca$.

If $a \leq b$, $a \cup b = b$; hence by (3), it follows that $ca \cup cb = cb$, which implies that $ca \leq cb$. In particular $a^2 \leq ab \leq b^2$.

Consider the two relations ζ and its dual ζ' defined in M as follows: $a\, \zeta\, b$ and $b\, \zeta' a$ if $a^2 \leq b \leq a$. Evidently ζ and ζ' satisfy (A), (B) and (A′), (B′) respectively. Now let $a\, \zeta\, b$ and $c \geq b$, then $(a \cup c)^2 \leq (a^2 \cup ac \cup ca \cup c^2) \leq b \cup c = c \leq c \cup a$. This proves that ζ is an H-relation. The relation ζ' satisfies (C′), for if $b\, \zeta'\, a$ and $c \leq a$ then $c^2 \leq a^2 \leq b$. Hence, since $c^2 \leq c$, $c^2 \leq b \cap c \leq c$, i. e., $(b \cap c)\zeta'\, c$. Thus ζ' is a dual H-property (One readily verifies that ζ' satisfies also E_1). Hence the upper ζ-radical and the lower ζ'-radical are defined in M.

In case M is the multiplicative lattice of the ideals of an associative ring S, the upper ζ-radical is known as the Baer's lower radical of S ([1]). It is readily seen that the theory of this radical is valid also in non-associative rings. Further properties of this radical will be discussed in Part III of this paper.

The lower ζ'-radical can be approached from a different point of view, due to the fact that the notion of $\bar{\zeta}'$-elements coincides with the idea of the

idempotent elements of M. Indeed, if $a^2 = a$ it is evident that a is a ζ'-element in M and conversely, since $a^2 \zeta' a$, the fact that a is a ζ'-element must yield $a^2 = a$. Thus we obtain by the dual of Theorem 1.2 that the lower ζ'-radical of M is the maximal idempotent element of M, and that the interval $[l(M,\zeta'), I]$ does not contain idempotent elements. This result has been obtained for ideals in rings and in semi-groups by J. Levitzki in [2].

HEBREW UNIVERSITY,
JERUSALEM, ISRAEL.

BIBLIOGRAPHY.

[1] R. Baer, "Radical ideals," *American Journal of Mathematics*, vol. 65 (1943), pp. 537-568.

[2] J. Levitzki, "On powers with transfinite exponents," (Hebrew) *Riveon Lematematica*, vol. 1 (1946), pp. 8-13.

[3] N. Jacobson, "The radical and semi simplicity for arbitrary rings," *American Journal of Mathematics*, vol. 67 (1945), pp. 300-320.

[4] B. Brown and N. H. McCoy, "Some theorems on groups with applications to ring theory," *Transactions of the American Mathematical Society*, vol. 69 (1950), pp. 302-311.

ON THE STRUCTURE OF MULTIPLY TRANSITIVE PERMUTATION GROUPS.* [1]

By T. C. HOLYOKE.

1. Introduction. Since Jordan [1] first considered the problem of transitive extension of finite transitive permutation groups [2], not much improvement on his work has been made in the way of general results. Zassenhaus [2] obtained results for certain groups of automorphisms of finite geometries. Witt [3, pp. 257-259] specialized Jordan's results to the case of multiply transitive groups, giving his results in such a way that they are still valid if infinite groups are admitted. Miller [4] showed that if a finite group of degree $n > 12$ does not contain the alternating group on its letters and is s-fold transitive, then $s < 3n^{\frac{1}{2}} - 2$. From this result it is possible to infer an upper limit of the number of successive transitive extensions which a given finite group can possess.

In this paper, the theory is extended to cover transitive extensions in general, the finite group appearing in the rôle of a special case for which certain simplifications of a calculational nature can be made. A general transitive extension theory, paralleling Schreier's extension theory, is given. Then the case of extension of a multiply transitive group is taken up in detail, in such a way as to emphasize the important rôle played by certain conjugates of the given group, not only insofar as the existence of transitive extensions is concerned, but also in the actual construction of them. Unlike the results of Jordan and Witt, the process given here is constructive in nature, reducing both the number of permutations to be investigated and the number of tests that need be applied to them, and leading eventually to the realization of the strong dependence of any multiply transitive group on certain associated groups which are exactly doubly and triply transitive.

2. Statement and general solution of the problem. Let there be given a permutation group G fixing a certain letter, say O, and transitive on the remaining letters (not necessarily finite or countable in number). We wish to find a permutation group H which is transitive on all the given letters and

* Received February 27, 1951; revised June 11, 1951.

[1] This paper is based on the author's doctoral dissertation at The Ohio State University.

in which G is the (largest) subgroup fixing the letter O. Such a group H has been referred to as a transitive extension of G [2].

For sections **2** and **3**, the following notation will be used: The letters displaced by G will be denoted by $\{x\}$. The subgroup of G fixing the letter $1 \varepsilon \{x\}$ will be denoted by F. For each letter x, a representative $u_x \varepsilon G$ will be prescribed which displaces 1 by x. This can be done because G is transitive. Let $u_1 = I$. Then $G = \Sigma_x F u_x$. A permutation $g \varepsilon G$ will be considered the same as

$$\begin{pmatrix} Fu_x \\ Fu_x g \end{pmatrix} \text{ or } \begin{pmatrix} u_x^{-1}F \\ g^{-1}u_x^{-1}F \end{pmatrix},$$

upon identification of letters x, Fu_x and $u_x^{-1}F$. The group G will be given as acting only on letters $\{x\}$, but will be considered the same as that group having the same permutations and fixing any other desired letters. H will denote a transitive extension of G, if one exists. The symbols f, g, and h are generic for elements of F, G, and H respectively, and the symbol u for the chosen coset representatives of F in G. Functions will be denoted by Greek letters, placed to the right of the elements on which they operate, composition of functions by $\circ$; transformation by a permutation p will be denoted by τ_p, and the composite function $\pi^{-1} \circ \gamma \circ \pi$ by γ^π, except that $\tau_p^{-1} \circ \gamma \circ \tau_p$ will be denoted by γ^p. Finally, $g_1(g_2\gamma)$ will be written as $g_1g_2\gamma$, and $(g_1\gamma) \cdot g_2$ as $g_1\gamma g_2$.

Suppose a transitive extension H exists. The following facts, well-known in the finite case, will be of use:

2.1 H contains an element a transposing 0 and 1, hence $a^2 \varepsilon F$ [1, p. 31]. If $\{x\}$ is finite, a^2 can be taken as I.

2.2 $a^{-1}Fa = F$. The proof here differs from that in the finite case [1, p. 32] in that $a^{-1}Fa$ could conceivably be a proper subgroup of F. It is not, however, since a^2 induces an automorphism of F.

2.3 For each $g \not\varepsilon F$ there exist g_1 and $g_2 \not\varepsilon F$ such that $aga = g_1 a g_2$ [1, p. 32].

2.4 $H = G + \Sigma_x G a u_x$, and every h has the form g or the form gau [cf. 3, p. 258].

2.5 $g_1 a g_2 = g_3 a g_4$ if and only if $g_3^{-1} g_1 \varepsilon F$ and $(g_3^{-1} g_2)\tau_a = g_4 g_2^{-1}$.

The following theorem serves to formalize the problem of transitive extension in terms of the internal structure of G.

2.6 THEOREM. *A necessary and sufficient condition for the existence of an extension H is that there exist an element q and an automorphism α*

of F, and two single valued mappings β and γ of $G-F$ into itself, satisfying the following relations for every g, g_1, and g_2 not in F and every f, f_1, and f_2 in F:

(a) $g\gamma$ is a u, $q\alpha = q$, $f\alpha \circ \alpha = f\alpha^2 = f\tau_q$;

(b) $\{f_1(f_1^{-1}\alpha g f_2)\beta\}^{-1} g\beta \in F$ and
$$(\{f_1(f_1^{-1}\alpha g f_2)\beta\}^{-1} g\beta)\alpha = (f_1^{-1}\alpha g f_2)\gamma f_2^{-1}\alpha(g\gamma)^{-1};$$

(c) $g\beta(g\gamma f)\beta \in F$ and $\{g\beta(g\gamma f)\beta\}\alpha = gqf\alpha\{(g\gamma f)\gamma\}^{-1}$;

(d) $\{(fg\beta)\beta\}^{-1}qf\alpha g \in F$ and $(\{(fg\beta)\beta\}^{-1}qf\alpha g)\alpha = \{fg\beta\}\gamma(g\gamma)$;

(e) $g_1(g_2\beta) \in F$ if and only if $g_1\gamma g_2 \in F$, in which case
$$g_1\beta q(g_1\gamma g_2)\alpha = q\{g_1 g_2\beta\}\alpha g_2\gamma;$$

(f) For $g_1g_2\beta$ not in F, $\{(g_1g_2\beta)\beta\}^{-1}g_1\beta(g_1\gamma g_2)\beta \in F$, and
$$(\{(g_1g_2\beta)\beta\}^{-1}g_1\beta(g_1\gamma g_2)\beta)\alpha = \{g_1g_2\beta\}\gamma g_2\gamma\{(g_1\gamma g_2)\gamma\}^{-1}.$$

Proof. Assume the existence of H. By 2.1 and 2.2, H contains an element a such that $a^2 \in F$ and $a^{-1}Fa = F$. Take α and q as τ_a and a^2, respectively. By 2.3, 2.4, and 2.5, for $g \notin F$, $aga = g'au$ (uniquely), with g' and u not in F. Define $g\beta = g'$ and $g\gamma = u$. Relations (b)-(f) now follow from 2.5 (except in the case of (e)) and the associativity of H, considering respectively elements of the following types: $a(f_1^{-1}\alpha g f_2)a$, $agafa$, $afaga$, ag_1ag_2a with $g_1g_2\beta \in F$, and ag_1ag_2a with $g_1g_2\beta \notin F$. Conversely, assume that q, α, β, and γ are given and satisfy the prescribed conditions. A suitable group H will be constructed. Consider all objects of types (g) and (g_1, g_2) for all g, g_1, and g_2 in G. Define the equivalence $\sim$: $(g) \sim (g')$ if and only if $g = g'$; $(g) \nsim (g_1, g_2)$; $(g_1, g_2) \sim (g_3, g_4)$ if and only if $g_3^{-1}g_1 \in F$ and $(g_3^{-1}g_1)\alpha = g_4g_2^{-1}$. Denote equivalence classes by $[g]$ and $[g_1, g_2]$. Let $\bar{H}$ denote the set of equivalence classes. Products are defined as follows:

$$[g_1][g_2] = [g_1g_2], \qquad [g_1][g_2, g_3] = [g_1g_2, g_3], \qquad [g_1, g_2][g_3] = [g_1, g_2g_3],$$
$$[g_1, g_2][g_3, g_4] = [g_1q(g_2g_3)\alpha g_4]$$

if $g_2g_3 \in F$, and $[g_1, g_2][g_3, g_4] = [g_1(g_2g_3)\beta, (g_2g_3)\gamma g_4]$ otherwise. That these products are well-defined follows directly from the definition of equivalence, except in the last two cases. For those cases, suppose $(g_1', g_2') \sim (g_1, g_2)$ and $(g_3', g_4') \sim (g_3, g_4)$. Then there exist f_1 and f_2 such that $g_1' = g_1f_1$, $g_2' = f_1^{-1}\alpha g_2$, $g_3' = g_3f_2$, and $g_4' = f_2^{-1}\alpha g_4$. Therefore g_2g_3 belongs to F if and only if $g_2'g_3'$ belongs to F, in which case

$$g_1'q(g_2'g_3')\alpha g_4' = g_1f_1qf_1^{-1}\alpha^2(g_2g_3)\alpha g_4 = g_1q(g_2g_3)\alpha g_4,$$

by (a). If g_2g_3 does not belong to F, then the equality of $[g_1, g_2][g_3, g_4]$ and $[g_1', g_2'][g_3', g_4']$ follows from (b) and the definition of equivalence.

The element $[I]$ is a right identity, and right inverses for elements $[g]$ and $[g_1, g_2]$ exist in $[g^{-1}]$ and $[g_2^{-1}, (g_1 q)^{-1}]$ respectively.

That products in $\bar{H}$ are associative follows directly from the definition of products and the associativity of G, except in the case of $[g_1, g_2]([g_3, g_4][g_5, g_6])$. For the various possibilities $g_2 g_3 \varepsilon F$ and $g_4 g_5 \varepsilon F$; $g_2 g_3 \not\varepsilon F$ and $g_4 g_5 \varepsilon F$; $g_2 g_3 \varepsilon F$ and $g_4 g_5 \not\varepsilon F$; $g_2 g_3 \not\varepsilon F$ and $g_4 g_5 \not\varepsilon F$ and $g_2 g_3 (g_4 g_5)\beta \varepsilon F$; and finally $g_4 g_5 \not\varepsilon F$ and $g_2 g_3 \not\varepsilon F$ and $g_2 g_3 (g_4 g_5)\beta \not\varepsilon F$, associativity follows from relations (a), (c), (d), (e), and (f) respectively. Hence $\bar{H}$ is a group. Define $\bar{a} = [I, I]$, $\bar{g} = [g]$, $\bar{G}$ the set of all $\bar{g}$, $\bar{F}$ the set of all $\bar{f}$. Then $\bar{G}$ is isomorphic to G, and $[g_1, g_2] = \bar{g}_1 \bar{a} \bar{g}_2$. Also, $\overline{a^{-1}} \bar{f} \bar{a} = [q(q\alpha)^{-1} f\alpha] = \overline{f\alpha}$, and $\bar{a}^2 = \bar{q}$. For $\bar{g} \not\varepsilon \bar{F}$, $\bar{a}\bar{g}\bar{a} = [g\beta, gy] = \overline{g\beta}\, \bar{\alpha}\, \overline{g\gamma}$. $\bar{H} = \bar{G} + \Sigma_x \bar{G}\bar{a}\bar{u}_x$, for, given $\bar{h} = [g_1, g_2]$, $\bar{h}$ can also be written in the form $[g, u]$, where $g_2 \varepsilon Fu$. Thus $h = \bar{g}\bar{a}\bar{u}$. The cosets are clearly distinct. It remains to be shown that $\bar{H}$ has a faithful permutation representation H on letters O and $\{x\}$ in which the image of $\bar{a}$ displaces O and the image of $\bar{g}$ is exactly the permutation g with the added fixed letter O. Such a representation is given by $\bar{h} \to \begin{pmatrix} \bar{G} \cdots & \bar{G}\bar{a}\bar{u} & \cdots \\ \bar{G}h \cdots & \bar{G}\bar{a}\bar{u}h & \cdots \end{pmatrix}$, under the correspondence $O \rightleftarrows \bar{G}$, $x \rightleftarrows \bar{G}\bar{a}\bar{u}_x$.

Here $\bar{a} \to \begin{pmatrix} \bar{G} & \bar{G}\bar{a} & \cdots \\ \bar{G}\bar{a} & \bar{G}\bar{a}^2 & \cdots \end{pmatrix} = \begin{pmatrix} 0 & 1 & \cdots \\ 1 & 0 & \cdots \end{pmatrix}$, and $\bar{g} \to \begin{pmatrix} \bar{G} \cdots & \bar{G}\bar{a}\bar{u} & \cdots \\ \bar{G} \cdots & \bar{G}\bar{a}\bar{u}\bar{g} & \cdots \end{pmatrix}$. But $g = \begin{pmatrix} Fu \\ Fug \end{pmatrix}$, and $\bar{G}\bar{a}\bar{u}\bar{g} = \bar{G}\bar{a}\bar{u}_x$ if and only if $Fug = Fu_x$. Thus H has the properties asserted in the theorem.

2.8 *Definition.* If q, α, β, and γ satisfy the conditions of Theorem 2.7, $(q, \alpha, \beta, \gamma)$ will be called an *embedding system.*

Notation. The "bar" notation of Theorem 2.7 will now be discarded. Let the mappings ϕ and ψ be defined as follows: $g\phi$ is the coset representative u for g, and $g\psi$ the element of F such that $g = g\psi g\phi$.

2.9 *Definition.* Two systems $(q, \alpha, \beta, \gamma)$ and $(q^*, \alpha^*, \beta^*, \gamma^*)$ are *similar* if there exist an element f and an automorphism π of G mapping F onto itself such that $g\gamma^* = (g\gamma^\pi f\alpha^\pi)\phi$ for all $g \not\varepsilon F$. Two similar systems for which π can be taken as the identity are *equivalent.*

2.10 *Definition.* Two transitive groups of permutations on the same set of letters are *permutation-isomorphic* if there exists a permutation of those letters transforming one onto the other.

2.11 Theorem. *Two transitive extensions of G are permutation-isomorphic if and only if the corresponding embedding systems are similar, and are identical if and only if the corresponding systems are equivalent,*

under the natural correspondence of letters Gau with Ga^*u and G with G, respectively.

Proof. Suppose H and H^* are permutation-isomorphic. Of the permutations transforming H onto H^*, at least one fixes the letters 0 and 1, since H^* is doubly transitive. Let p be such a permutation. Then $G\tau_p = G$ and $F\tau_p = F$. Also, $a\tau_p \varepsilon Fa^*$, so that $a^* = fp^{-1}ap$. Now

$$\begin{aligned} g\beta^* a^* g\gamma^* &= a^* g a^* = fp^{-1}apgfp^{-1}ap = fp^{-1}ag\tau_{p^{-1}}af\tau_{p^{-1}} \circ \alpha p \\ &= fg\beta^p a\tau_p(g\gamma^p f\alpha^p)\psi(g\gamma^p f\alpha^p)\phi = [fg\beta^p(g\gamma^p f\alpha^p)\psi \circ (\alpha^{-1})^p f^{-1}]a^*(g\gamma^p f\alpha^p)\phi, \end{aligned}$$

so that $g\gamma^*$ has the desired form, with $\pi = \tau_p$.

Conversely, suppose two systems and an element f are given such that $g\gamma^* = (g\gamma^\pi f\alpha^\pi)\phi$, where π is an automorphism of G mapping F onto itself. Recall that in H^* the letter Ga^*u, and in H the letter Gau, correspond to the letter Fu as given in G. In the following discussion, all representatives u are different from the identity. There exists a permutation p fixing letters 0 and 1 such that $\pi = \tau_p$ [3, p. 259], in particular, $p = \begin{pmatrix} Fu \\ Fu\pi \end{pmatrix}$. To show that $p^{-1}Hp = H^*$, it will suffice now, by 2.4, to show that $a^* = fa\tau_p$.

$$\begin{aligned} a\tau_p = (01)\begin{pmatrix} (Gau)p \\ (Gau\gamma)p \end{pmatrix} &= (01)\begin{pmatrix} (Fu)p \\ (Fu\gamma)p \end{pmatrix} = (01)\begin{pmatrix} Fu\pi \\ Fu\gamma \circ \pi \end{pmatrix} \\ &= (01)\begin{pmatrix} Gau\pi \\ Gau\gamma \circ \pi \end{pmatrix} = (01)\begin{pmatrix} Ga^*u' \\ Ga^*u'\gamma^\pi \end{pmatrix}. \end{aligned}$$

Hence

$$\begin{aligned} fa\tau_p = a\tau_p f\alpha^\pi = (01)\begin{pmatrix} Ga^*u \\ Ga^*u\gamma^\pi \end{pmatrix} \cdot (0)(1)\begin{pmatrix} Ga^*u\gamma^\pi \\ Ga^*u\gamma^\pi f\alpha^\pi \end{pmatrix} &= (01)\begin{pmatrix} Ga^*u \\ Ga^*u\gamma^\pi f\alpha^\pi \end{pmatrix} \\ = (01)\begin{pmatrix} Ga^*u \\ Ga^*(u\gamma^\pi f\alpha^\pi)\phi \end{pmatrix} &= (01)\begin{pmatrix} Ga^*u \\ Ga^*u\gamma^* \end{pmatrix} = a^*. \end{aligned}$$

If π is the identity automorphism, then p can be taken as the identity permutation, and $fa = a^*$, hence H and H^* are the same.

2.12 Corollary. The relations *similar* and *equivalent* of 2.9 are equivalences.

The importance of the concept of similarity of systems is brought out in section 4.

The connection between the mappings β and γ and the permutation a can be partially expressed in terms of permutations. Define the permutations b and c (on the cosets of F in G as letters) by the following: $(F)b = (F)c = F$; $(gF)b = g\beta F$ and $(Fg)c = Fg\gamma$ for $g \notin F$. If b and c are considered as permutations of the letters of G, they fix the letter 1, and

$b^{-1} = c = (01) \cdot a$. Furthermore, $G_{T_0} \not\subset G$ unless G contains all finite even permutations of its letters.

2.13 *Remark.* G has no transitive extension if it does not contain all finite even permutations of its letters and if its normalizer in the group of all permutations of letters $\{x\} - 1$ contains the normalizer of F on those letters.

3. Some groups having no transitive extensions. The following theorems indicate two infinite classes of finite simply transitive groups for which transitive extensions do not exist.

3.1 THEOREM. *Let G be a transitive permutation group on letters $1, 2, \cdots, n, n+1, n+2, \cdots, n+n$ such that the subgroup F fixing the letter 1 also fixes $n+1$, and has two multiply transitive constituents on the sets $2, \cdots, n$ and $n+2, \cdots, n+n$ respectively, so constituted that in each element f, the fixed letters are congruent $(\operatorname{mod} n)$ in pairs, and congruent letters are displaced by congruent letters; or let G be permutation-isomorphic to such a group. Then G has no transitive extension.*

Proof. Permutation-isomorphic groups have or fail to have transitive extensions together; so the group G can be taken as described. F contains an elements f of order two, having an even number of transpositions, say $2j$. Then $2j \leqq n - 1$. Suppose an extension H exists. H contains an element a interchanging 0 and 1, and conjugate to f [1, p. 31], therefore having $2j - 1$ transpositions among letters $2, 3, \cdots, 2n$. Now $a^{-1}Fa = F$ (2.2), hence a fixes the letter $n+1$ and either permutes the letters of each transitive set among themselves, or must interchange the sets entirely. The latter case is impossible, for a would have $n-1$ transpositions among $2, 3, \cdots, 2n$. Therefore it may be supposed that a takes each transitive set into itself. Then a has the transposition (ij) among letters $2, \cdots, n$ if and only if it has the transposition $(n+i, n+j)$, since different permutations induce different automorphisms of a multiply transitive group. This contradicts the fact that $2j - 1$ is odd. Hence no such group H exists.

3.2 COROLLARY. *If the symmetric group of degree $n > 4$ is represented as a group G of permutations on the cosets of one of its alternating subgroups of degree $n - 1$, then G has no transitive extension.*

3.3 THEOREM.[3] *If G is a transitive group $(\neq I)$ of odd order and of degree $n \equiv 1 \pmod 4$, then G has no transitive extension.*

[3] The author is indebted to M. Hall for this theorem.

Proof. Suppose an extension H exists. The order of H is divisible by 2 but not by 4; hence H has a Sylow subgroup of order 2. Since a subgroup of order 2 is in the center of its normalizer, H contains a normal subgroup H^* of index 2 in H [5, p. 327]. G, being maximal in H, is not contained in H^*. Therefore $G \cap H^*$ has index 2 in G, yielding a contradiction.

4. Multiply transitive groups. The notation will now be changed in accordance with the following: for $k > 1$, let G_k represent a k-fold transitive group on letters $\{x\}$, $0, 1, \cdots, k-1$, not containing an element consisting of a cycle of length 2 or 3; and for $i = 0, 1, \cdots, k-1$, let G_i be the subgroup of G_{i+1} fixing the letter i. The letter added in the attempt to find transitive extensions of G_k will be called k. Let $G_k^{(\rho)}$, for variable index ρ, represent all the distinct transitive extensions of G_{k-1} other than G_k.

4.1 *Remark.* In G_i $(i = 2, \cdots, k)$, there exists an element a_i transposing letters $i-1$ and $i-2$, permuting letters $\{x\}$ among themselves, and fixing the rest [cf. 1, p. 32, and 3, p. 257]. Let elements similar to a_k be chosen also in each $G_k^{(\rho)}$ and designated by $a_k^{(\rho)}$. All these elements are permutable with G_0 [3, p. 257].

4.2 *Definition.* Let $a_{k+1}^{(\rho)} = a_k^{(\rho)\tau_{a_k}} \cdot (k, k-1, k-2)$.

4.3 *Remark.* If adjunction of various elements $a_{k+1}^{(\rho)}$ to G_k produces transitive extensions of G_k, then all such extensions are distinct.

Proof. If $a_{k+1}^{(\rho)}$ and $a_{k+1}^{(\rho')}$ yield the same extension, then

$$a_{k+1}^{(\rho)} \varepsilon\, G_0 a_{k+1}^{(\rho')},$$

from which it follows that $a_k^{(\rho)} \varepsilon\, G_0 a_k^{(\rho')}$, and $G_k^{(\rho)} = G_k^{(\rho')}$, i. e. $\rho = \rho'$.

4.4 THEOREM. *In order for a particular $a_{k+1}^{(\rho)}$ to yield a transitive extension of G_k, it is necessary and sufficient that $(a_k^{(\rho)}a_k)^3 \varepsilon\, G_0$ and $a_{k+1}^{(\rho)}G_{k-1}a_{k+1}^{(\rho)} = G_{k-1}$. For $k > 2$, these conditions are equivalent to the two conditions $(a_k^{(\rho)}a_k)^3 \varepsilon\, G_0$* and *$(a_{k+1}^{(\rho)}a_{k-1})^3 \varepsilon\, G_0$ (in Theorem 4.11, it is shown that for $k > 3$, the single condition $(a_{k+1}^{(\rho)}a_{k-1})^2 \varepsilon\, G_0$ is sufficient).*

Proof. Since $(a_k a_{k+1}^{(\rho)})^3 = (a_k^{(\rho)}a_k)^3$, this theorem follows from a theorem of Jordan [1, p. 33] and Witt [3, p. 257].

4.5 THEOREM. *Each transitive extension of G_k can be obtained by adjoining to it one of the elements $a_{k+1}^{(\rho)}$, in which case the corresponding $G_k^{(\rho)}$ (uniquely determined, as follows from 4.3) is permutation-isomorphic to G_k, the element $a_{k+1}^{(\rho)} \cdot (k, k-1)$ transforming G_{k-1} into itself and a_k into some element of the coset $G_0 a_k^{(\rho)}$.*

Proof. Let G_{k+1} be a transitive extension of G_k, a_{k+1} an element of G_{k+1} which interchanges k and $k-1$, and fixes $k-2, \cdots, 0$. Define $a_k' = a_k\tau_{a_{k+1}} \cdot (k, k-1, k-2)$. Then $a_k' \notin G_k$, otherwise $(k, k-1, k-2)$ would belong to G_{k+1}, in which case G_k would also contain a cycle of length 3, since G_{k+1} is triply transitive. G_k and $G_{k-1} \cup \{a_k'\}$ are permutation-isomorphic, since the element $a_{k+1} \cdot (k, k-1)$ transforms G_{k-1} into itself and a_k into a_k'. $G_{k-1} \cup \{a_k'\}$ is an extension of G_{k-1} and is therefore one of the $G_k^{(\rho)}$. Now $a_k^{(\rho)} = fa_k'$ for some element $f \varepsilon G_0$, and

$$
\begin{aligned}
a_{k+1}^{(\rho)} &= a_k^{(\rho)}\tau_{a_k} \cdot (k, k-1, k-2) \\
&= a_k^{-1} f a_{k+1} a_k a_{k+1} \cdot (k, k-1, k-2) \cdot a_k \cdot (k, k-1, k-2) \\
&= a_k^{-1} f a_{k+1}^{-1} a_k a_{k+1} a_k \; \varepsilon \; G_{k+1},
\end{aligned}
$$

hence $G_{k+1} = G_k \cup \{a_{k+1}^{(\rho)}\}$ (2.4), and $a_{k+1}^{(\rho)} = f_1 a_{k+1}$ for some element $f_1 \varepsilon G_0$. Since all the a's are permutable with G_0, the final assertion of the theorem follows.

4.6 Corollary. *If the number of distinct extensions of G_{k-1} which are permutation-isomorphic to G_k is a finite number t, then the number of extensions of G_k is less than t; in particular, if a finite simply transitive group G of degree $n > 3$ has at most t permutation-isomorphic extensions, then G cannot be extended to a $(t+1)$-fold transitive group of degree $n+t+1$.*

4.7 *Definition.* If every $a_{k+1}^{(\rho)}$ produces a transitive extension of G_k, it will be said that G_k *can be extended in every possible way.*

4.8 Lemma. *If, for $k > 3$, G_k and G_k' are two extensions of G_{k-1}, then there exist permutations s_1, s_2, s_3, and s_4 on letters $\{x\}$ such that* $G_{k-2}^{(i)} = G_{k-3} \cup \{s_i \cdot (k-3, k-4)\}$ $(i = 1, 2, 3, 4)$ are transitive extensions of G_{k-3}, $G_{k-1}^{(j)} = G_{k-2}^{(1)} \cup \{s_j\tau_{s_1} \cdot (k-2, k-3)\}$ $(j = 2, 3, 4)$ are transitive extensions of G_{k-2} $(= G_{k-2}^{(1)})$, and $G_k^{(m)} = G_{k-1}^{(2)} \cup \{s_m\tau_{s_2} \circ \tau_{s_1} \cdot (k-1, k-2)\}$ $(m = 3, 4)$ are the two given extensions G_k and G_k', respectively of G_{k-1} $(= G_{k-1}^{(2)})$.

Proof. This follows from successive application of 4.2-4.5.

Let τ_i represent τ_{s_i} throughout the remainder of this paper.

4.9 Lemma. $s_i^3 \equiv (s_2 s_m)^3 \equiv I \pmod{G_0}$,

$$s_m\tau_2 \circ \tau_1 \equiv s_m\tau_2, \text{ and } s_4\tau_3 \circ \tau_2 \circ \tau_1 \equiv s_4\tau_3 \circ \tau_2,$$

$$(i = 1, 2, 3, 4;\ m = 3, 4).$$

Proof. $s_i^2 \equiv I$, by (2.1). $(s_2 s_m)^3 \equiv I$, since, by 4.4,

$$\{s_2\tau_1 \cdot (k-2, k-3) \cdot s_m\tau_1 \cdot (k-2, k-3)\}^3 \equiv I.$$

By Theorem 4.4, $I \equiv (s_m\tau_2 \circ \tau_1 s_1)^2 \equiv s_m\tau_2 \circ \tau_1 s_m \tau_2$, from which it follows that $s_m\tau_2 \circ \tau_1 \equiv s_m^{-1}\tau_2 \equiv s_m\tau_2$. Finally,

$$\begin{aligned} s_4\tau_3 \circ \tau_2 &\equiv s_3\tau_2 s_4 \tau_2 s_3 \tau_2 \\ &\equiv s_3\tau_2 \circ \tau_1 s_4 \tau_2 \circ \tau_1 s_3 \tau_2 \circ \tau_1 \equiv (s_3 s_4 s_3)\tau_2 \circ \tau_1 \\ &\equiv s_4\tau_3 \circ \tau_2\tau_1. \end{aligned}$$

4.10 Lemma. *$(s_4\tau_3 \circ \tau_2 s_2 \tau_1)^2 \equiv I$, if and only if $(s_4\tau_3 \circ \tau_2 s_2)^2 \equiv I$, and if these relations hold, then $(s_4\tau_3 \circ \tau_2 s_3 \tau_2)^3 \equiv I$.*

Proof. Lemma 4.9 will be used here repeatedly. Assume

$$(s_4\tau_3 \circ \tau_2 s_2 \tau_1)^2 \equiv I.$$

Then

$$I \equiv (s_4\tau_3 \circ \tau_2 \circ \tau_1 s_2 \tau_1)^2 \equiv (s_4\tau_3 \circ \tau_2 s_2)^2 \tau_1,$$

whence $(s_4\tau_3 \circ \tau_2 s_2)^2 \equiv I$. Assume that $(s_4\tau_3 \circ \tau_2 s_2)^2 \equiv I$. Then

$$(s_4\tau_3 \circ \tau_2 s_2 \tau_1)^2 \equiv (s_4\tau_3 \circ \tau_2 \circ \tau_1 s_2 \tau_1)^2 \equiv (s_4\tau_3 \circ \tau_2 s_2)^2 \tau_1 \equiv I.$$

Also $s_4\tau_3 \circ \tau_2 \equiv s_4^{-1}\tau_3 \equiv s_4\tau_3$, and

$$(s_4\tau_3 \circ \tau_2 s_3 \tau_2)^3 \equiv (s_4\tau_3 s_3 \tau_2)^3 \equiv (s_4\tau_3 s_2 \tau_3)^3 \equiv (s_4 s_2)^3 \tau_3 \equiv I.$$

4.11 Theorem. *In order for $a_{k+1}^{(\rho)}$ (cf. 4.2 and 4.4) to yield a transitive extension of G_k when $k > 3$, it is sufficient that $(a_{k+1}^{(\rho)} a_{k-1})^2 \in G_0$.*

Proof. Let the groups G_k and G_k' of 4.8-4.10 correspond to the groups G_k and $G_k^{(\rho)}$ of this theorem, respectively, $s_3\tau_2 \circ \tau_1 \cdot (k-1, k-2)$ and $s_4\tau_2 \circ \tau_1 \cdot (k-1, k-2)$ to a_k and $a_k^{(\rho)}$, and $s_2\tau_1 \cdot (k-2, k-3)$ to a_{k-1}. Then $a_{k+1}^{(\rho)} = s_4\tau_3 \circ \tau_2 \circ \tau_1 \cdot (k, k-1)$, and the theorem is seen to follow from 4.9, 4.10, and 4.4.

4.12 Theorem. *If, for $k > 3$, every $(k-1)$-fold transitive extension of G_{k-3} can be extended in every possible way (cf. 4.7), then all the extensions of G_{k-1} are permutation-isomorphic and can themselves be extended in every possible way.*

Proof. If G_k and $G_k^{(\rho)}$ are two extensions of G_{k-1}, it is required to show that $a_{k+1}^{(\rho)}$ yields a transitive extension of G_k. Let the same correspondence be set up as in the proof of the previous theorem. Then the following groups are $(k-1)$-fold transitive extension of G_{k-3}:

$$\{G_{k-3}, s_1(k-3, k-4), s_2\tau_1\cdot(k-2, k-3)\} \ (=G_{k-1}),$$
$$\{G_{k-3}, s_1(k-3, k-4), s_3\tau_1(k-2, k-3)\},$$

and

$$\{G_{k-3}, s_1(k-3, k-4), s_4\tau_1(k-2, k-3)\}.$$

Since $G_{k-3}\tau_2{}^{s_1} = G_{k-3}$ (Theorem 4. 4), it follows that

$$\{G_{k-3}, s_1(k-3, k-4), s_3\tau_1(k-2, k-3)\}\tau_2{}^{s_1}$$
$$= \{G_{k-3}, s_2(k-3, k-4), s_3\tau_2(k-2, k-3)\}$$

and

$$\{G_{k-3}, s_1(k-3, k-4), s_4\tau_1(k-2, k-3)\}\tau_2{}^{s_1}$$
$$= \{G_{k-3}, s_2(k-3, k-4), s_4\tau_2(k-2, k-3)\}$$

are also $(k-1)$-fold transitive extensions of G_{k-3}. The assumption that each such group can be extended in every possible way leads to the conclusion that $I \equiv (s_3{}^{-1}\tau_2 s_4 \tau_2 s_3 \tau_2 s_2)^2 \equiv (s_4\tau_3 \circ \tau_2 s_2)^2 \pmod{G_0}$; and the desired result now follows from 4. 10 and 4. 11.

4. 13 *Remark.* It is easy to show that each permutation of the letters $\{x\}$ which, with the adjunction of a suitable transposition, is used to obtain an extension of G_1 having degree of transitivity greater than three, has already been used similarly in obtaining some extension of G_1 which is exactly triply transitive. This fact seems to lend weight to the conjecture that there are only a finite number of triply transitive groups which have transitive extensions.

NORTHWESTERN UNIVERSITY.

BIBLIOGRAPHY.

[1] C. Jordan, *Traité des Substitutions et des Équations Algébriques*, Paris, 1870.

[2] H. Zassenhaus, "Über transitive Erweiterung gewisser Gruppen aus Automorphismen endlicher mehrdimensionaler Geometrien," *Mathematische Annalen*, vol. 111 (1935), pp. 748-759.

[3] E. Witt, "Die 5-fach transitiven Gruppen von Mathieu," *Abhandlung aus dem Mathematischen Seminar der Hamburgischen Universitat* 12 (1938), pp. 256-264.

[4] G. A. Miller, "Limits of the degree of transitivity of substitution groups," *Bulletin of the American Mathematical Society* (2), vol. 22 (1915), pp. 68-71.

[5] W. Burnside, *Theory of Groups of Finite Order*, 2nd ed., Cambridge, 1911.

ON THE DETERMINATION OF MULTIPLY CONNECTED DOMAINS OF AN ELASTIC PLANE BODY, BOUNDED BY FREE BOUNDARIES WITH CONSTANT TANGENTIAL STRESSES.*[1]

By C. Arf.

1. Formulation of the problem.

B. Consider the closed curves L of the z plane which are boundaries of multiply connected domains C of multiplicity m, such that there exists a state of stress of C which satisfies the following conditions.

B$_1$. The stress tensor $(\sigma_x, \sigma_y, \tau)$ satisfies the condition $\sigma_x + \sigma_y = 4\alpha$ with a given constant α throughout C and on L.

B$_2$. The stress tensor $(\sigma_x, \sigma_y, \tau)$ has determined finite or infinite values everywhere in C and on the boundary L, with a possible exception at $z = x + iy = \infty$. More precisely the analytic function $\Gamma(z) = (\sigma_x - \sigma_y - 2i\tau)/4\alpha$ has no singularities other than poles in C and on the boundary L except at the point $z = \infty$

B$_3$. The boundary L is a free boundary for the state of stress $(\sigma_x, \sigma_y, \tau)$.

B$_4$. $\Gamma(z)$ has only a finite number of poles in C and on L. Besides these we shall require that the following geometrical conditions are satisfied.

B$_5$. The domain C does not contain the point $z = \infty$. But the point $z = \infty$ might be a point of the boundary L. For a variable point z of the domain which tends toward ∞ continuously, $\arg z$ has limits which are the numbers $\phi_1, \phi_2, \cdots, \phi_n$ $(0 \leq \phi_1 < \phi_2 < \cdots < \phi_n < 2\pi)$.

B$_6$. To every ϕ_j is associated a set of separated intervals

$$(a_1^{(j)}, b_1^{(j)}), (a_2^{(j)}, b_2^{(j)}), \cdots, (a_{\nu_j}^{(j)}, b_{\nu_j}^{(j)}),$$

which are described by the expression $\lim_{z\to\infty} \Im z \exp(-i\phi_j)$, where $z \in C$ is supposed to go to ∞ continuously and $\lim_{z\to\infty} \arg z = \phi_j$.

* Received February 20, 1951.

[1] I wish to thank Dr. R. C. Roberts for helping me prepare the English version of this paper.

B_7. L has only a finite number of points of inflection and a finite number of cuspidal points.

It is clear that these conditions require that the curve L have the form shown in (fig. 1) and that the isolated exterior forces which are applied to the elastic plane body $(C+L)$ are like those shown on that figure.

The purpose of this paper is to determine the domains C and the stress tensors σ_x, σ_y, τ which satisfy the conditions above.

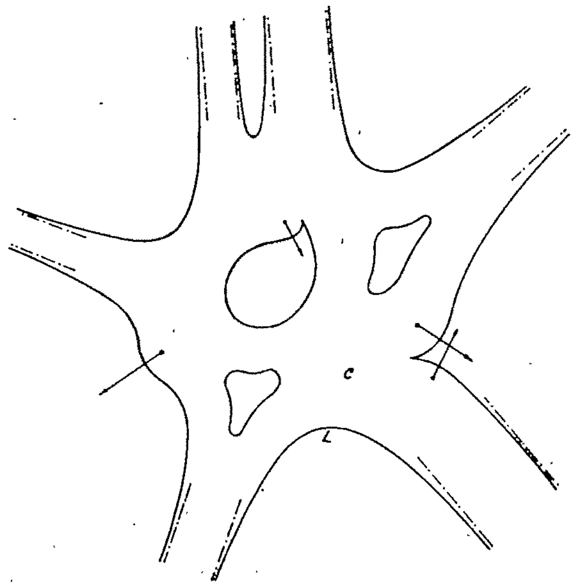


FIG. 1.

2. In our former publications[2] we have shown that a state of stress for which $\sigma_x+\sigma_y=4\alpha$ is a constant has free boundaries if and only if it is defined in the following manner except for the case where the stresses are constants throughout the plane.

Let $\chi(t)$ be an analytic function which takes real values on some parts of the real axis of the T-plane. We put

$$z=\Phi(\zeta)-i\int\chi(i(\zeta-1)/(\zeta+1))\,d\zeta/\zeta^2 \tag{2.1}$$

The function $z=\Phi(\zeta)$ defines a conformal mapping (locally) of the ζ-plane onto the z-plane. At the point $z=\Phi(\zeta)$ we consider the tensor

[2] C. Arf, "Sur les frontières libres à tensions constantes d'un milieu élastique plan en équilibre," *Notes communiquées au VI^e congrès international de mécanique appliquée*, 1946.

$$\begin{aligned}\sigma_x &= 2\alpha(1+\Re\zeta^2),\\ \sigma_y &= 2\alpha(1-\Re\zeta^2),\\ \tau &= -2\alpha\Im\zeta^2,\end{aligned}\tag{2.2}$$

which defines a state of elastic stress of the z-plane. The point

$$z = \Phi((1-it)/(1+it))\tag{2.3}$$

describes a free boundary of this state of stress, when t varies on the part of the real axis of the plane where $\chi(t)$ is real. If ϕ denotes the angle of the tangent to the boundary with the x axis, the center of curvature of L at the corresponding point z of the boundary will be

$$z + i\chi(tg\tfrac{1}{2}\phi)e^{i\phi},\tag{2.4}$$

and we will have

$$z = \Phi((1-itg\tfrac{1}{2}\phi)/(1+itg\tfrac{1}{2}\phi)) = \Phi(e^{-i\phi}), \qquad \zeta = e^{-i\phi}\tag{2.5}$$

3. We consider now the function

$$Z = \zeta^2 = (\sigma_x - \sigma_y - 2i\tau)/4\alpha = \Gamma(z)\tag{3.1}$$

which transforms conformally our domain C and its boundary L into a Riemann surface $\mathfrak{L}$ and its boundary k respectively. The point $Z = \Gamma(z)$ tends toward the point $Z = \exp(-2i\phi_l)$ when z tends toward $z = \infty$ continuously under the condition $\lim\limits_{z\to\infty} \arg z = \phi_l$. The set $\mathfrak{L} + k = E$, in which we include the points $\exp(-2i\phi_l)$, is then a closed domain of a certain Riemann surface over the Z plane. We consider now the symmetric surface $\mathfrak{L}^s$ of $\mathfrak{L}$ with respect to the boundary k of E. The set $\mathfrak{L}^s + \mathfrak{L} + k = R$ will then be a closed Riemann surface. Let us consider a connected part L_1 of L. According to our assumptions, L has only a finite number of cuspidal points and a finite number of points of inflection. Therefore the variation of the angle ϕ of the tangent to L, when the point z describes the connected part L_1, will be such that $Z = \exp(-2i\phi)$ will draw a finite number of arcs of finite length of the unit circle $|Z| = 1$. The boundary k will then be constituted by m such closed arcs. If we now take into account that $Z = \Gamma(z)$ can take the value $Z = \infty$ only a finite number of times, or that E has only a finite number of points at infinity. We can conclude that E and therefore R has only a finite number of sheets. *R is therefore an algebraic Riemann surface of genus $m-1$ and has only a finite number of points of ramification.*

Let M_{Z_0} be a point of E different from the images of the point $z = \infty$.

Let $t = (Z - Z_0)^{1/\mu}$ (or $t = Z^{-1/\mu}$ if $Z_0 = \infty$) be the local uniformization parameter of R at this point M_{Z_0}. We consider the inverse function $z = \Lambda(Z)$ of $Z = \Gamma(z)$ at this point M_{Z_0}. Because of the fact that the mapping $z = \Lambda(Z)$ of E into $C + L$ is conformal, we shall have

$$(3.2) \qquad \Lambda(Z) = a_0 + a_1 t + \cdots + a_n t^n + \cdots \qquad \text{with } a_1 \neq 0,$$

and therefore

$$(3.3) \qquad \Lambda'(Z) = 1/\mu t^{-\mu+1}(a_1 + 2a_2 t + \cdots + n a_n t^{n-1} + \cdots) \text{ for } Z_0 \text{ finite},$$

$$(3.3') \qquad \Lambda'(Z) = \mathrm{v}/\mu t^{\mu+1}(a_1 + 2a_2 t + \cdots + n a_n t^{n-1} + \cdots) \text{ for } Z_0 = \infty.$$

From this we can deduce the following results:

An ordinary point of E which does not correspond to a point at infinity on L is an ordinary point of $\Lambda'(Z)$.

A point of ramification of order μ of E which does not correspond to a point at infinity on L is a pole of order $\mu - 1$ of $\Lambda'(Z)$, if its affix is finite.

A point M_∞ of ramification of order μ on E is a zero of order $\mu + 1$ of $\Lambda'(Z)$.

The function $\Lambda(Z)$ is obviously equal to $\Lambda(Z) = \Phi(Z^{\frac{1}{2}})$. We have therefore

$$(3.4) \qquad \Lambda'(Z) = \Phi'(Z^{\frac{1}{2}})/2Z^{\frac{1}{2}} = i\chi(i(Z^{\frac{1}{2}} - 1)/(Z^{\frac{1}{2}} + 1))/2ZZ^{\frac{1}{2}},$$

or

$$(3.5) \qquad -\chi(iZ^{\frac{1}{2}} - 1)/(Z^{\frac{1}{2}} + 1)) = 2iZZ^{\frac{1}{2}}\Lambda'(Z).$$

Thus the function $iZZ^{\frac{1}{2}}\Lambda'(Z)$ has to be real on the boundary k of E. We can therefore extend this function on $\mathfrak{L}^s$ by the formula

$$(3.6) \qquad iZZ^{\frac{1}{2}}\Lambda'(Z) = \overline{iZ^*Z^{*\frac{1}{2}}\Lambda'(Z^*)}$$

where Z^* is the symmetrical point of Z with respect to k. One can write this as follows:

$$(3.7) \qquad \Lambda'(Z) = -\overline{Z^{*3}\Lambda'(Z^*)}$$

This relation enables us to extend the function $\Lambda'(Z)$ on $\mathfrak{L}^s$. Our results about the singularities of $\Lambda'(Z)$ on $\mathfrak{L}$ will then have the following consequences on $\mathfrak{L}^s$.

A point of ramification of order μ on $\mathfrak{L}^s$ is a pole of order $\mu - 1$ of $\Lambda'(Z)$, if this point has a finite affix different from zero.

A point of ramification of order μ on $\mathfrak{R}^s$ whose affix is zero, is a pole of order $2\mu - 1$ of $\Lambda'(Z)$.

A point at infinity on $\mathfrak{R}^s$ is a zero of order $2\mu + 1$ of $\Lambda'(Z)$, if this point has an order of ramification μ on $\mathfrak{R}^s$.

We can also show in exactly the same way as we have done in our former work [3] that a *point P of R which corresponds to a point at infinity on L is a pole of order μ for $\Lambda'(Z)$, if this point P is a point of ramification of order μ on R.*

In the following parts of the paper we shall study only the special case $m = 2$. The case $m > 2$ will be studied in a forthcoming paper.

4. Let us denote by L_1 the exterior boundary of our domain C, and by L_2 its interior boundary. L_2 will have no point at infinity.

For $m = 2$ the genus of R is 1, and we can represent Z and $\Lambda'(Z)$ as elliptic functions $H(T)$ and $R(T)$ of a variable $T = t + it'$.

Let ω_1, ω_2 be the fundamental periods of these functions, and let us consider the corresponding σ-function of Weierstrass. It follows from the fact that R is symmetrical with respect to k that to each zero a of order μ of $Z = H(T)$ corresponds a pole b of order μ of $H(T)$. The function $H(T)$ has then the following form

$$H(T) = A \prod_{j=1}^{\nu} (\sigma(T - a_j)/\sigma(T - b_j))^{\mu_j}, \tag{4.1}$$

where the points $a_1, a_2, \cdots, a_\nu$; $b_1, b_2, \cdots, b_\nu$ are different from each other and satisfy the condition

$$\sum_{j=1}^{\nu} \mu_j a_j = \sum_{j=1}^{\nu} \mu_j b_j. \tag{4.2}$$

We consider now a point M on R and the symmetrical point M^s with respect to k. Let T and T^s be the corresponding points of the T-plane.

$$T \to \bar{T}^s \tag{4.3}$$

is then a conformal mapping of the finite T-plane on itself. This mapping is therefore of the form

$$\bar{T}^s = \alpha T + \beta \qquad \alpha, \beta \text{ constants.} \tag{4.4}$$

[3] C. Arf, "Sur la détermination des états d'équilibres d'un milieu élastique plan admettant des frontières libres à tensions constantes," *Revue de la Faculté des Sciences de l'Université d'Istanbul*, tome XII, Serie A, 1947.

From $(M^s)^s = M$ we deduce $(T^s)^s \equiv T \bmod(\omega_1, \omega_2)$, which means

$$(4.5)\qquad \begin{aligned} &\bar\alpha(\alpha T+\beta)+\bar\beta \equiv T && \bmod(\omega_1,\omega_2),\\ &|\alpha| = 1, \qquad \bar\alpha\beta+\bar\beta\equiv 0 && \bmod(\omega_1,\omega_2). \end{aligned}$$

By a preliminary substitution of the form $T \to aT + b$ we can make $\alpha = 1$ and β real.[4] We can then either choose as one of our periods ω_1, ω_2 (ω_1 for example) a real number, and we have $\beta = \omega_1/2$ or $\beta = 0$. Besides this, to every set of congruent points $T + m_1\omega_1 + m_2\omega_2$ must correspond a set $T^s + m_1'\omega_1 + m_2'\omega_2$ of congruent points. This means that we must have $\bar T + m_1\omega_1 + m_2\bar\omega_2 + \beta = \bar T + \beta + m_1'\omega_1 + m_2'\omega_2$. Therefore the conjugates of periods must be periods too. The fundamental periods are then either of the form $\omega_1 = \omega$, $\omega_2 = i\omega'$, or of the form $\omega_1 = 2\omega$, $\omega_2 = \omega + i\omega'$ (ω, ω' real). If $\beta = \omega_1/2$, the images of k in the T-plane are the lines $T = \bar T + \omega_1/2 + m_1\omega_1 + m_2\omega_2$, or

$$(4.6)\qquad 2\Im T = m_2\omega', \qquad m_1 + \tfrac12 = 0, \qquad \text{for } \omega_1 = \omega,\ \omega_2 = i\omega',$$

$$(4.7)\qquad 2\Im T = m_2\omega', \qquad m_2 + 2m_1 + 1 = 0 \qquad \text{for } \omega_1 = 2\omega,\ \omega_2 = \omega + i\omega'.$$

The equations (4.6) have no solutions, while the lines which are solutions of (4.7) are deduced from one of them by the translations $n\omega_2$. In the case $\beta = \omega_1/2$ the boundary k will have therefore either no point at all or consist only of one piece.

We must have therefore $\beta = 0$. The images of k in the T-planes are then the lines

$$(4.8)\qquad 2\Im T = m_2\omega', \qquad m_1 = 0, \qquad \text{for } \omega_1 = \omega,\ \omega_2 = i\omega',$$

$$(4.9)\qquad 2\Im T = m_2\omega', \qquad m_2 + 2m_1 = 0, \qquad \text{for } \omega_1 = 2\omega,\ \omega_2 = \omega + i\omega'.$$

The lines which are solutions of (4.9) are obtained from one of them by the translations $m\omega_2$. The boundary k will consist in that case of a curve of one piece only. *We shall therefore have to consider only the case*

$$(4.10)\qquad \omega_1 = \omega, \qquad \omega_2 = i\omega' \qquad T^s = \bar T\,. \qquad (\omega, \omega' \text{ real})$$

The fundamental periods being of this form, the functions

$$\sigma(T) = T\Pi'\{(1 - T/(m\omega + im'\omega'))\exp(T/(m\omega + im'\omega') + T^2/2(m\omega + im'\omega')^2)\},$$

$$\zeta(T) = 1/T + \Sigma'(1/(T - m\omega - im'\omega') + 1/(m\omega + im'\omega') + T/(m\omega + im'\omega')^2),$$

[4] Such a substitution carries the coefficients α and β into $a\alpha/\bar a$, $\beta + (\bar a b - a\alpha\bar b)/\bar a$.

will be real on the real axis, and we shall have $\sigma(\bar{T}) = \sigma(T)$, $\zeta(\bar{T}) = \zeta(T)$. The points of affix zero and the points at infinity on R being symmetrical with respect to k, the corresponding points a_j and b_j of the T plane will have to satisfy the condition

$$b_j \equiv \bar{a}_j, \qquad \mathrm{mod}(\omega, i\omega'). \tag{4.11}$$

We consider now the functions $R(T) = \Lambda'(Z) = \Lambda'(H(T))$ and $F(T) = \Lambda(Z) = \Lambda(H(T))$. Since the functions $R(T)$ and $H(T)$ are elliptic of periods $\omega, i\omega'$, the function

$$F'(T) = \Lambda'(Z) \cdot H'(T) = R(T)H'(T) \tag{4.12}$$

is also elliptic of periods $\omega, i\omega'$. The function $\Lambda'(Z)$ has to satisfy the condition

$$\Lambda'(Z) = -\overline{\Lambda'(Z)z^3},$$

at every point Z of k (3.7). The function $F'(T)$ will have therefore to satisfy the condition

$$F'(T)/H'(T) = -\overline{H(T)}^3\,\overline{F'(T)}/\overline{H'(T)} \tag{4.13}$$

for all values of T for which $\Im T = 0$ or $\omega'/2$. But for these values of T we have

$$Z\bar{Z} = H(T) \cdot \overline{H(T)} = 1, \qquad H'(T)\overline{H(T)} + \overline{H'(T)}H(T) = 0.$$

The condition (4.13) can therefore be written as follows:

$$H(T) = \overline{F'(T)}/F'(T), \qquad \text{for } \Im T = 0, \omega'/2. \tag{4.14}$$

If we put $\underline{F}'(T) = \overline{F'(\bar{T})}$, we can deduce from (4.14) and from the fact that $F'(T)$ is an elliptic function of periods $(\omega, i\omega')$ that

$$H(T) = \underline{F}'(T)/F'(T), \tag{4.15}$$

for all values of T. Reciprocally one can deduce from (4.15) the relations (4.14), $H(T) \cdot \overline{H(T)} = 1$ and (4.13) for $\Im T = 0$ or $\omega'/2$. The function $z = \Lambda(Z) = F(T)$ maps one of the strips

$$0 < \Im T < \omega'/2, \qquad -\omega'/2 < \Im T < 0$$

on our domain C. But by a substitution of the form $T \to aT + b$ ($a = \pm 1$, $b = 0, i\omega'/2$) we can make this function map the strip $0 < \Im T < \omega'/2$ on the domain C and the line $\Im T = 0$ on the exterior boundary L_1 of C. This mapping must obviously be conformal in that strip $0 < \Im T < \omega'/2$. *The*

zeros and the poles of $F'(T)$ are therefore in the strips which can be obtained from $-\omega'/2 \leq \Im T \leq 0$ by translations of magnitudes $m\omega'$

Let us consider a point T_0 on $\Im T = 0$ or $\Im T = \omega'/2$. If the image $F(T_0) = \Lambda(H(T_0))$ of this point on the boundary L is different from ∞ and from a cuspidal point of L, then the point $Z_0 = H(T_0)$ of k is a pole of order $\mu - 1$ of $\Lambda'(Z)$, where μ is the order of ramification of this point on R. The point T_0 is then a zero of order $\mu - 1$ of $H'(T)$, and the function $F'(T) = \Lambda'(H(T)) \cdot H'(T)$ will be regular at the point T_0. If the image $F(T_0)$ is a point at infinity on L_1 (in which case $\Im T_0 = 0$), the point $Z_0 = H(T_0)$ on k is a pole of order μ of $\Lambda'(Z) = \Lambda'(H(T))$, and the function $F'(T) = \Lambda'(H(T)) \cdot H'(T)$ has a simple pole at T_0.

The function $F'(T)$ has therefore no singularities on $\Im T = \omega'/2$, and its only singularities on $\Im T = 0$ are poles of order 1.

The zeros of $F'(T)$ on the lines $\Im T = 0$, $\Im T = \omega'/2$ have also to be of order 1. In fact, if T_0 is a zero of $F'(T)$ on one of these lines, then the image of the neighborhood of T_0 in $0 < \Im T < \omega'/2$ by the mapping $z = F(T)$ will have more than one sheet unless $F''(T_0) \neq 0$. Under this condition the neighborhood of $F(T_0)$ in C will have to be of one of the following three forms [5]

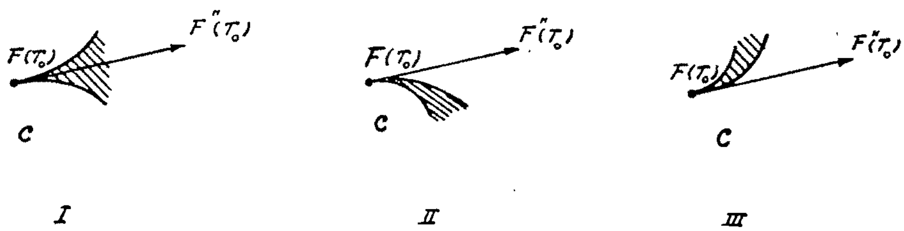


FIG. 2.

But it can be easily shown that the cuspidal point $F(T_0)$ of L has one of these three forms if and only if

$$(4.16) \quad \left.\begin{array}{ll} \Im(F^{(\nu)}(T_0)/F''(T_0)) > 0 & \text{for } \Im T_0 = 0 \\ \Im(F^{(\nu)}(T_0)/F''(T_0)) < 0 & \text{for } \Im T_0 = \omega'/2 \end{array}\right\} \text{for odd values of } \nu,$$

where $F^{(\nu)}(T)$ is the lowest order derivative of $F(T)$ such that

$$\Im(F^{(\nu)}(T_0)/F''(T_0)) \neq 0.$$

[5] C. Arf, "Sur la détermination des états d'équilibres d'un milieu élastique plan admettant des frontières libres à tensions constantes," ***Revue de la Faculté des Sciences de l'Université d'Istanbul***, tome XII, Serie A, 1947.

Finally, the function $z = F(T) = \int F'(T)\,dT$ has to be chosen in such a way that the domain $0 \leq \Im T \leq \omega'/2$, $0 \leq \Re T < \omega$ is mapped on a doubly connected closed domain of the z plane. For this it is necessary to have

$$F(T+\omega) - F(T) = \int_0^\omega F'(T+t)\,dt = 0 \tag{4.17}$$

for all values of T with $0 < \Im T < \omega'/2$. In order to make this condition explicit we consider the following representation of $F'(T)$:

$$\begin{aligned} F'(T) = c + \sum_{j=1}^{n} \{ & A_j \zeta(T-\beta_j) + A'_j \zeta'(T-\beta_j) \\ & + \cdots + A_j^{(\mu_j-1)} \zeta^{(\mu_j-1)}(T-\beta_j) \}, \end{aligned} \tag{4.18}$$

where

$$\begin{aligned} A_j(T-\beta_j)^{-1} - A'_j(T-\beta_j)^{-2} + 2A''_j(T-\beta_j)^{-3} \\ - \cdots + (-1)^{\mu_j-1}(\mu_j-1)!A_j^{(\mu_j-1)}(T-\beta_j)^{-\mu_j} \end{aligned} \tag{4.19}$$

is the principal part of $F'(T)$ at the pole β_j, c is a suitable constant, and where the coefficients A_j have to satisfy the condition

$$\sum_{j=1}^{n} A_j = 0. \tag{4.20}$$

We put now this representation of $F'(T)$ in (4.17), and we obtain the following condition for the coefficients A_j, A'_j, and the constant c:

$$\begin{aligned} 0 = \int_0^\omega F'(T+t)\,dt = c\omega + \sum_{j=1}^{n} \{ & A_j \int_0^\omega \zeta(T+t-\beta_j)\,dt \\ & + A'_j(\zeta(T+\omega-\beta_j) - \zeta(T-\beta_j))\}. \end{aligned} \tag{4.21}$$

Using the known relations

$$\zeta(T-\beta_j+\omega) - \zeta(T-\beta_j) = 2\zeta(\omega/2) = \eta,$$

$$\begin{aligned} \int_0^\omega \zeta(T-\beta_j+t)\,dt &= \log(\sigma(T-\beta_j+\omega)/\sigma(T-\beta_j)), \\ &= \eta(T-\beta_j+\omega/2) + i\pi, \end{aligned}$$

we can write this condition as follows

$$c\omega - \eta\Big(\sum_{j=1}^{n} A_j\beta_j - \sum_{j=1}^{n} A'_j\Big) = 0. \tag{4.22}$$

We consider again a pole $T = \beta$ of $F'(T)$ on the line $\Im T = 0$. As we have noticed before, this pole is of order one, and in the neighborhood of β we have

$$(4.23) \qquad F(T) = A\int\zeta(T-\beta)dT + R_1(T)$$
$$= A\log\sigma(T-\beta) + R_1(T) = A\log(T-\beta) + R(T),$$

where $R_1(T)$, $R(T)$ are regular in the neighborhood of $T=\beta$. If one follows the variation of this function in the neighborhood of β in the strip $0 \leq \Im T \leq \omega'/2$, one sees that the function $F(T)$ maps such a neighborhood on a domain like the one shown in figure 3.

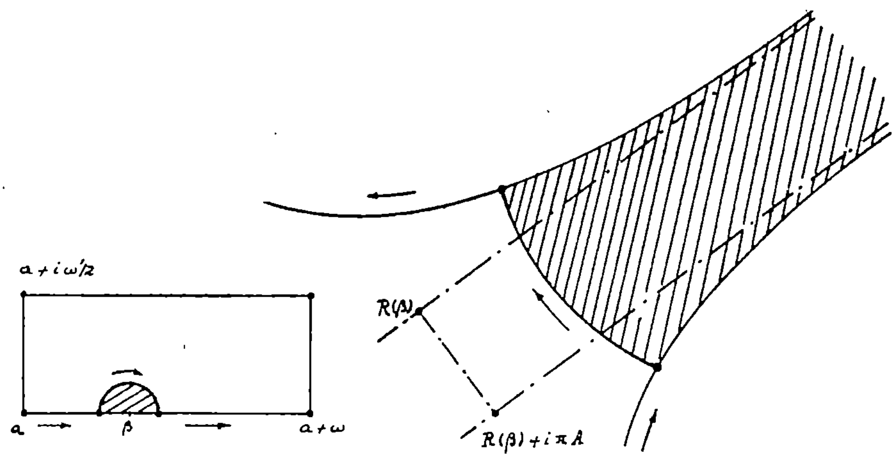


FIG. 3.

The neighborhood of $T=\beta$ corresponds to an infinite strip of c whose direction is given by the coefficient $-A$ of the principal part of $F'(T)$ at $T=\beta$. The thickness of this strip is obviously $|A|\pi$. If therefore $\beta_1, \beta_2, \cdots, \beta_h$ is the increasing sequence of the poles of $F'(T)$ in $0 \leqq T < \omega$, and $A_1 = -f_1\exp(i\phi_1)$, $A_2 = -f_2\exp(i\phi_2), \cdots$, $A_h = -f_h\exp(i\phi_h)$ are the corresponding residues of $F'(T)$, we must have

$$0 \leqq \phi_2 - \phi_1 \leqq \phi_3 - \phi_1 \leqq \cdots \leqq \phi_h - \phi_1 \leqq 2\pi,$$

in order that the corresponding strips of c do not overlap each other.

Besides this we must also require that $F'(T)$ be such that none of the closed curves

$$z = F(i(\omega'/2 - 0) + t), \qquad 0 \leq t < \omega,$$

$$z = F(+i0 + t), \qquad 0 \leq t < \omega,$$

cuts itself.

Conversely let a function

$$F'(T) = \omega^{-1}\eta\Big(\sum_{j=1}^{k}(A_j\gamma_j - A'_j) - \sum_{j=1}^{h}\beta_j f_j \exp(i\phi_j)\Big) - \sum_{j=1}^{h} f_j \exp(i\phi_j)\zeta(T-\beta_j)$$
$$+ \sum_{j=1}^{k}\{A_j\zeta(T-\gamma_j) + A'_j\zeta'(T-\gamma_j) + \cdots + A_j^{(\mu_j-1)}\zeta^{(\mu_j-1)}(T-\gamma_j)\}$$

be given which satisfies the following six conditions:

I. $$\sum_{j=1}^{k} A_j - \sum_{j=1}^{h} f_j \exp(i\phi_j) = 0,$$

II. $$0 \leq \beta_1 < \beta_2 < \cdots < \beta_h < \omega,$$

III. $$f_j > 0, \qquad 0 \leq \phi_2 - \phi_1 < \phi_3 - \phi_1 < \cdots < \phi_h - \phi_1 \leq 2\pi$$

IV. $$-\omega'/2 < \Im\gamma_j < 0, \qquad 0 \leq \Re\gamma_j < \omega.$$

V. *The function $F'(T)$ has no zero in $0 < \Im T < \omega'/2$, and its zeros α_l on the lines $\Im T = 0$, $\Im T = \omega'/2$ are all simple and satisfy the condition*

$$\Im(F^{(\nu)}(\alpha_l)/F''(\alpha_l)) > 0 \quad \textit{for } \Im\alpha_l = 0 \textit{ and } \nu \textit{ odd},$$
$$\Im(F^{(\nu)}(\alpha_l)/F''(\alpha_l)) < 0 \quad \textit{for } \Im\alpha_l = \omega'/2 \textit{ and } \nu \textit{ odd},$$

where $F^{(\nu)}(T)$ is the first derivative of $F(T)$ for which $\Im(F^{(\nu)}(\alpha_l)/F''(\alpha_l)) \neq 0$.

VI. *None of the curves*

$$z = F(+i0 + t), \qquad 0 \leq t < \omega,$$
$$z = F(i(\omega'/2 - 0) + t), \qquad 0 \leq t < \omega,$$

cuts itself.

The function

$$z = F(T) = \omega^{-1}\eta\Big(\sum_{j=1}^{k}(A_j\gamma_j - A'_j) - \sum_{j=1}^{n} f_j\beta_j \exp(i\phi_j)\Big)T \tag{4.24}$$
$$- \sum_{j=1}^{h} f_j \exp(i\phi_j)\log(\sigma(T-\beta_j))$$
$$+ \sum_{j=1}^{k}\{A_j \log(\sigma(T-\gamma_j)) + A'_j\zeta(T-\gamma_j)$$
$$+ \cdots + A_j^{(\mu_j-1)}\zeta^{(\mu_j-2)}(T-\gamma_j)\}$$

maps the domain $0 \leq \Im T \leq \omega'/2$, $0 \leq \Re T < \omega$ on a doubly connected simple domain C which satisfies the conditions B_5, B_6, B_7, and the distribution of stress given by the formula

$$(4.25) \qquad x+iy=z=F(T),$$

$$\sigma_x=2\alpha(1+\Re(\underline{F}'(T)/F'(T))), \qquad \sigma_y=2\alpha(1-\Re(\underline{F}'(T)/F'(T))),$$

$$\tau=-2\alpha\Im(\underline{F}'(T)/F'(T)), \qquad \underline{F}'(T)=\overline{F'(\bar{T})}$$

satisfies the conditions B_1, B_2, B_3, B_4.

5. In this section we shall compute the following quantities.

a) The resultant $X+iY$ of the forces which are exerted on the boundary (b) of a small body containing the point $z_j=F(\bar{\gamma}_j)$ where $\underline{F}'(T)/F'(T) = (\sigma_x-\sigma_y-2i\tau)/4\alpha$ becomes infinite,

b) The resulting momentum of these forces with respect to the point $F(\bar{\gamma}_j)$,

c) The displacement $u+iv$ which corresponds to the state of stress given by (4.25).

$$\text{a)} \qquad X+iY=\oint_{(b)}(\tau dx-\sigma_x dy)+i\oint_{(b)}(\sigma_y dx-\tau dy)$$

$$=2\alpha\oint_{(b)}-\Im(\underline{F}'(T)/F'(T))dx-(1+\Re(\underline{F}'(T)/F'(T)))dy,$$

$$+2\alpha i\oint_{(b)}(1-\Re(\underline{F}'(T)/F'(T)))dx+\Im(\underline{F}'(T)/F'(T))dy,$$

$$(5.1) \qquad X+iY=2\alpha i\oint_{(b)}dz-\overline{(\underline{F}'(T)/F'(T))dz}=-2\alpha i\oint_{(b_T)}F'(\bar{T})d\bar{T},$$

where (b_T) designates the image of (b) in the T-plane. Since

$$A_j(T-\gamma_j)^{-1}-A'_j(T-\gamma_j)^{-2}+\cdots+(-1)^{\mu_j-1}(\mu_j-1)!A_j^{(\mu_j-1)}(T-\gamma_j)^{-\mu_j}$$

is the principal part of $F'(T)$ at the point γ_j, the value of this integral is $X_j+iY_j=-2\alpha i\overline{(2\pi i A_j)}$, so that

$$(5.2) \qquad X_j+iY_j=-4\pi\alpha A_j.$$

The thickness of the strip which corresponds to the pole β_j being πf_j, the resultant of the forces exerted at infinity on that strip is $-4\pi\alpha f_j\exp(i\phi_j)$. We remark therefore that the condition

$$(\mathrm{I}) \qquad \sum_{j=1}^{k} f_j\exp(i\phi_j)-\sum_{j=1}^{k} A_j=0,$$

which expresses the fact that the function $F'(T)$ is elliptic, also expresses the fact that the resultant of the external forces acting on (C), is zero.

b) The resulting moment of the forces which act on the boundary (b) is given by

$$M_j = \oint_{(b)} (x - \Re F(\bar\gamma_j))(\sigma_y dx - \tau dy) - (y - \Im F(\bar\gamma_j))(\tau dx - \sigma_x dy),$$

$$= 2\alpha \oint_{(b)} (1 - \Re(\underline{F}'(T)/F'(T)))(x - \Re F(\bar\gamma_j))dx + \Im(F'(T)/F'(T)(x - \Re F(\bar\gamma_j))dy$$
$$+ \Im(\underline{F}'(T)/F'(T))(y - \Im F(\bar\gamma_j)dx + (1 + \Re(\underline{F}'(T)/F'(T)))(y - \Im F(\bar\gamma_j))dy,$$

$$= 2\alpha \oint_{(b)} \tfrac{1}{2}d \mid z - F(\bar\gamma_j)\mid^2 - \Re(\underline{F}'(T)/F'(T))\Re((z - F(\bar\gamma_j))dz)$$
$$+ \Im(\underline{F}'(T)/F'(T))\Im((z - F(\bar\gamma_j))dz),$$

$$= 2\alpha \oint_{(b)} \tfrac{1}{2}d \mid z - F(\bar\gamma_j)\mid^2 - \Re[(\underline{F}'(T)/F'(T))(z - F(\bar\gamma_j))dz],$$

$$= 2\alpha \oint_{(b_T)} - \Re[\underline{F}'(T)(F(T) - F(\bar\gamma_j))dT],$$

$$= 4\alpha\pi\Im[\bar A'_j F'(\bar\gamma_j) - \bar A''_j F''(\bar\gamma_j) + \cdots + (-1)^{\mu_j}\bar A_j^{(\mu_j-1)}F^{(\mu_j-1)}(\bar\gamma_j)],$$

and the moment with respect to the origin is given by

$$M_j^0 = M_j - 4\pi\alpha\Im\bar A_j F(\bar\gamma_j).$$

The moment $\mathfrak{M}_j^0$ of the forces which are acting on the strip which corresponds to the pole β_j of $F'(T)$ is easy to compute. One finds

$$\mathfrak{M}_j^0 = -4\pi\alpha[\Im(F(T) + f_j \exp(i\phi_j)\log\sigma(T - \beta_j))_{T=\beta_j} - \tfrac{1}{2}\pi f_j^2].$$

If we compute now the sum

$$\sum_{j=1}^{h} \mathfrak{M}_j^0 - \sum_{j=1}^{k} M_j^0,$$

we obtain zero.

c) From the usual form

$$\begin{Bmatrix} \sigma_x & \tau \\ \tau & \sigma_y \end{Bmatrix} = \begin{Bmatrix} \mu u_x + \frac{1}{2}\lambda(u_x + v_y) & \frac{1}{2}\mu(u_y + v_x) \\ \frac{1}{2}\mu(u_y + v_x) & \mu v_y + \frac{1}{2}\lambda(u_x + v_y) \end{Bmatrix}$$

of Hooke's law one can deduce easily the following expression of the displacement vector which corresponds to the distribution of stress

$$\sigma_x = 2\alpha(1 + \Re(\underline{F}'(T)/F'(T))), \qquad \sigma_y = 2\alpha(1 - \Re\underline{F}'(T)/F'(T)),$$
$$\tau = -2\alpha\Im(\underline{F}'(T)/F'(T));$$
$$u + iv = (2\alpha/\mu)(z2\mu/(\lambda+\mu) + \int \underline{F}'(T)dT).$$

This displacement is obviously uniform in our domain C if and only if $A_1 = A_2 = \cdots = A_k = 0$, *and*

$$0 = \int_0^\omega \underline{F}'(T+t)dt \equiv \eta(-\sum_{j=1}^{k} \bar{A}'_j - \sum_{j=1}^{h} f_j \exp(-i\phi_j)\beta_j$$
$$+ \eta(\sum_{j=1}^{k} A'_j + \sum_{j=1}^{h} f_j \exp(-i\phi_j)\beta_j)$$

which is in fact identically satisfied.

6. In this section we shall consider the special case where the quantity $\underline{F}'(T)/F'(T) = (\sigma_x - \sigma_y - 2i\tau)/4\alpha$ is finite everywhere in C. In that case the function $\underline{F}'(T)$ can have no pole in the strip $0 < \Im T \leq \omega'/2$. $F'(T)$ is therefore of the form

$$F'(T) = -\sum_{j=1}^{h} f_j \exp(i\phi_j)\,(\zeta(T-\beta_j) + \eta\beta_j/\omega). \tag{6.1}$$

The function $F'(T)$ *must not have all its zeros on the lines* $\Im T = n\omega'$, $\Im T = \omega'/2 + m\omega'$; for in this case the function $H(T) = \underline{F}'(T)/F'(T)$ would be a constant. Indeed, if

$$F'(T) = A\prod_{l=1}^{h}(\sigma(\tau-\alpha_l)/\sigma(T-\beta_l)) \quad \text{with} \ \sum_{l=1}^{h}\alpha_l - \sum_{l=1}^{n}\beta_l = 0, \tag{6.2}$$

where the points α_l are on the lines $\Im T = n\omega'/2$, then $\bar{\alpha}_l = \alpha_l + n_l i\omega'$, with $\sum_{l=1}^{h} n_l = 0$, and

$$(6.3)\qquad \begin{aligned} \underline{F}'(T) &= \bar{A}\prod_{l=1}^{h}(\sigma(T-\alpha_l)/\sigma(T-\beta_l))\exp(-\sum_{l=1}^{h} n_l\eta'(T-\alpha_l - in_l\omega'/2)), \\ \underline{F}'(T) &= (\bar{A}/A)F'(T)\exp(\sum_{l=1}^{h} n_l\eta'(\alpha_l + in_l\omega'/2)). \end{aligned}$$

Now we are going to show that the function $F'(T)$ *must have no zero at all on the lines* $\Im T = n\omega'/2$.

Let us suppose that α_1, for example, is on one of these lines $\Im T = n\omega'/2$; α_1 is then a simple zero of $F'(T)$, and according to V. has to satisfy the condition

$$(6.4)\quad \begin{aligned} &F''(\alpha_1) \neq 0 \qquad \Im(F'''(\alpha_1)/F''(\alpha_1)) \geqq 0 \quad \text{for } n \text{ even,} \\ &F''(\alpha_1) \neq 0 \qquad \Im(F'''(\alpha_1)/F''(\alpha_1)) \leqq 0 \quad \text{for } n \text{ odd.} \end{aligned}$$

We have

$$(6.5)\qquad F''(T) = F'(T)\Big(\sum_{l=1}^{h} (\sigma'(T-\alpha_l)/\sigma(T-\alpha_l)) - \sum_{l=1}^{h} (\sigma'(T-\beta_l)/\sigma(T-\beta_l))\Big),$$

$$(6.6)\qquad \begin{aligned} F'''(T) = {} & F''(T)\Big(\sum_{l=1}^{h} (\sigma'(T-\alpha_l)/\sigma(T-\alpha_l)) - \sum_{l=1}^{h} (\sigma'(T-\beta_l)/\sigma(T-\beta_l))\Big) \\ & + F'(T)\Big(\sum_{l=1}^{h} (\sigma'(T-\alpha_l)/\sigma(T-\alpha_l))' - \sum_{l=1}^{h}(\sigma'(T-\beta_l)/\sigma(T-\beta_l))'\Big), \end{aligned}$$

and

$$(6.7)\quad \begin{aligned} F'''(T)/F''(T) = {} & F''(T)/F'(T) \\ & + \Big(\sum_{l=1}^{h} (\sigma'(T-\alpha_l)/\sigma(T-\alpha_l))' - \sum_{l=1}^{h}(\sigma'(T-\beta_l)/\sigma(T-\beta_l))'\Big)F'(T)/F''(T). \end{aligned}$$

From this formula we get

$$(6.8)\quad \begin{aligned} & F'''(\alpha_1)/F''(\alpha_1) \\ & = [F''(T)/F'(T) + (\sigma'(T-\alpha_1)/\sigma(T-\alpha_1))'F'(T)/F''(T)]_{T=\alpha_1} \end{aligned}$$

But from (6.5) we can deduce the following expression of $F''(T)/F'(T)$ in the neighborhood of $T=\alpha_1$:

$$(6.9)\qquad F''(T)/F'(T) = 1/\sigma(T-\alpha_1) + A_0 + A_1\sigma(T-\alpha_1) + \cdots$$

with

$$(6.10)\qquad A_0 = \sum_{l=2}^{h} (\sigma'(\alpha_1-\alpha_l)/\sigma(\alpha_1-\alpha_l)) - \sum_{l=1}^{h} (\sigma'(\alpha_1-\beta_l)/\sigma(\alpha_1-\beta_l)).$$

We have therefore

$$(6.11)\quad \begin{aligned} F'''(\alpha_1)/F''(\alpha_1) = {} & [A_0 + 1/\sigma(T-\alpha_1) \\ & + (\sigma'(T-\alpha_1)/\sigma(T-\alpha_1))'\sigma(T-\alpha_1)(1-A_0\sigma(T-\alpha_1)+\cdots)]_{T=\alpha_1}. \end{aligned}$$

We have also

$$(6.12)\qquad \sigma'(T-\alpha_1)/\sigma(T-\alpha_1) = 1/\sigma(T-\alpha_1) - \tfrac{1}{3}c_2\sigma^3(T-\alpha_1) + \cdots,$$

$$(6.13)\quad (\sigma'(T-\alpha_1)/\sigma(T-\alpha_1))' = -1/\sigma^2(T-\alpha_1) + \tfrac{2}{3}c_2\sigma^2(T-\alpha_1) + \cdots,$$

and therefore

$$(6.14) \qquad F'''(\alpha_1)/F''(\alpha_1) = 2A_0 = 2\Big(\sum_{l=2}^{h} \zeta(\alpha_1 - \alpha_l) + \sum_{l=1}^{h} \zeta(\alpha_1 - \beta_l)\Big),$$

$$(6.14^1) \qquad \Im(F'''(\alpha_1)/F''(\alpha_1)) = 2\Im\Big(\sum_{l=2}^{h} \zeta(\alpha_1 - \alpha_l) - \sum_{l=1}^{h} \zeta(\alpha_1 - \beta_l)\Big).$$

We have seen that the zeros $\alpha_1, \alpha_2, \cdots, \alpha_h$ of $F'(T)$ are congruent to some points $\alpha'_1, \alpha'_2, \cdots, \alpha'_h$ of the strip $-\omega'/2 \leq \Im T \leq 0$, and that at least one of these points α'_j is in the strip $-\omega'/2 < \Im T < 0$. Using this and the relations $\sum_{l=1}^{h} \alpha_l - \sum_{l=1}^{h} \beta_l = 0$, $\Im\beta_l = 0$, we can write $\alpha_l = \alpha'_l + in_l\omega'$, with

$$\Im \sum_{l=1}^{h} \alpha_l = \sum_{l=1}^{h} \Im\alpha'_l + \omega' \sum_{l=1}^{h} n_l = 0,$$

and

$$\Im(F'''(\alpha_1)/F''(\alpha_1)) = 2\Im\Big[\sum_{l=2}^{h} (\zeta(\alpha'_1 - \alpha'_l) - 2(n_l - n_1)\zeta(i\omega'/2)) - \sum_{l=1}^{h} (\zeta(\alpha'_1 - \beta_l) + 2n_1\zeta(i\omega'/2))\Big],$$

$$\Im(F'''(\alpha_1)/F''(\alpha_1)) = 2\Im\Big[\sum_{l=2}^{h} (\zeta(\alpha'_1 - \alpha'_l) - 2((\alpha'_1 - \alpha'_l)/i\omega')\zeta(i\omega'/2)) - \sum_{l=1}^{h} (\zeta(\alpha'_1 - \beta_l) - 2((\alpha'_1 - \beta_l)/i\omega')\zeta(i\omega'/2)\Big],$$

$$(6.15) \qquad \Im(F'''(\alpha_1)/F''(\alpha_1)) = (-4/\omega')\Re\Big[\sum_{l=2}^{h} (\zeta(\alpha'_1 - \alpha'_l)i\omega'/2 - (\alpha'_1 - \alpha'_l)\zeta(i\omega'/2)) - \sum_{l=1}^{h} (\zeta(\alpha'_1 - \beta_l)i\omega'/2 - (\alpha'_1 - \beta_l)\zeta(i\omega'/2)\Big].$$

We have now

$$(6.16) \qquad \begin{aligned} &\Re(\zeta(T)i\omega'/2 - T\zeta(i\omega'/2)) \geq 0, \quad \text{for } n\omega' \leq \Im T \leq n\omega' + \omega'/2, \\ &\Re(\zeta(T)i\omega'/2 - T\zeta(i\omega'/2)) \leq 0, \quad \text{for } n\omega' + \omega'/2 \leq \Im T \leq (n+1)\omega', \end{aligned}$$

where the equalities hold if and only if $\Im T = m\omega'/2$.[6] In (6.15) the points $\alpha'_1 - \beta_l$ are, according to our hypothesis, all on the line $\Im T = 0$ or $\Im T = -\omega'/2$, and the points $\alpha'_1 - \alpha'_l$ are in the strip

$$0 \leq \Im T \leq \omega'/2, \qquad \text{for } \Im\alpha'_1 = 0,$$

$$-\omega'/2 \leq \Im T \leq 0, \qquad \text{for } \Im\alpha'_1 = -\omega'/2,$$

[6] See Appendix.

at least one of them being inside of the corresponding strip. From this and from (6.15) we obtain the inequalities

$$\Im(F'''(\alpha_1)/F''(\alpha_1)) < 0 \qquad \text{for } \Im\alpha'_1 = 0$$

$$\Im(F'''(\alpha_1)/F''(\alpha_1)) > 0 \qquad \text{for } \Im\alpha'_1 = -\omega'/2$$

which contradict the condition (6.4).

The function $F'(T)$ must have therefore all its zeros inside the strips $n'\omega + \omega'/2 < \Im T < (n+1)\omega'$ and the curve L has no cuspidal point.

We will show now that *the curve L has no point of inflexion.*

Let us consider a point $z = F(t + i\epsilon\omega'/2)$ of L (t real and $\epsilon = 0$ or 1). The angle of the tangent to the boundary at this point, with the x axis being ϕ, we have

$$\underline{F}'(T)/F'(T) = H(T) = \exp(-2i\phi), \qquad T = t + i\epsilon\omega'/2,$$

and therefore

$$\begin{aligned} -2i\phi &= \log \underline{F}'(T) - \log F'(T) \\ &= \log \bar{A} - \log A + \sum_{l=1}^{h} (\log \sigma(T - \bar{\alpha}_l) - \log \sigma(T - \alpha_l)). \end{aligned}$$

It is then enough to show that the derivative of this expression does not vanish for real values of $t = T - i\epsilon\omega'/2$.

$$-2i\phi_t = \sum_{l=1}^{h} (\zeta(t + i\epsilon\omega'/2 - \bar{\alpha}_l) - \zeta(t + i\epsilon\omega'/2 - \alpha_l)),$$

$$\begin{aligned} -2\phi_t = \sum_{l=1}^{h} \Im(\zeta(t + i\epsilon\omega'/2 - \bar{\alpha}'_l) + 2n_l\zeta(\omega' i/2) \\ - \zeta(t + i\epsilon\omega'/2 - \alpha'_l) + 2n_l\zeta(\omega' i/2)), \end{aligned}$$

$$\begin{aligned} (6.17) \quad \phi_t = [&\sum_{l=1}^{h} \Re(\zeta(t + i\omega'\epsilon/2 - \bar{\alpha}'_l)\omega' i/2 - (t + i\omega'\epsilon/2 - \bar{\alpha}'_l)\zeta(\omega' i/2)) \\ &- \sum_{l=1}^{h} \Re(\zeta(t + i\omega'\epsilon/2 - \alpha'_l)\omega' i/2 - (t + i\omega'\epsilon/2 - \alpha'_l)\zeta(\omega' i/2))]1/\omega'. \end{aligned}$$

The points α'_l being all in the strip $-\omega'/2 < \Im T < 0$, the points $t + i\omega'\epsilon/2 - \bar{\alpha}'_l$ are all in the strip $-\omega'/2 < \Im T < 0$ for $\epsilon = 0$ and $0 < \Im T < \omega'/2$ for $\epsilon = 1$, while the points $t + i\omega'\epsilon/2 - \alpha'_l$ are all in

$0 < \Im T < \omega'/2$ for $\epsilon = 0$ and $\omega'/2 < \Im T < \omega'$ for $\epsilon = 1$. We obtain then by the inequalities (6.16)

$$\text{(6.18)} \qquad \phi_t < 0 \quad \text{for } \epsilon = 0; \qquad \phi_t > 0 \quad \text{for } \epsilon = 1.$$

Let us follow now the variation of ϕ when the point $z = F(T)$ describes one of the curves L_1, L_2. We have

$$\text{(6.19)} \quad \phi = (1/\omega') \int \sum_{l=1}^{h} \Re(\zeta(t + i\omega'\epsilon/2 - \bar{\alpha}'_l)\omega' i/2 - (t + i\omega'\epsilon/2 - \bar{\alpha}'_l)\zeta(\omega' i/2))\, dt - \sum_{l=1}^{h} \Re(\zeta(t + i\omega'\epsilon/2 - \alpha'_l)\omega' i/2 - (t + i\omega'\epsilon/2 - \alpha'_l)\zeta(\omega' i/2))dt.$$

The variation of ϕ along these lines is therefore

$$\text{(6.20)} \quad \delta\phi = (1/\omega')\Re \sum_{l=1}^{h} \Big(\int_{\tau^*_l}^{\tau^*_l+\omega} (\zeta(\tau)\omega' i/2 - \tau\zeta(\omega' i/2))d\tau - \int_{\tau_l}^{\tau_l+\omega} (\zeta(\tau)\omega' i/2 - \tau\zeta(\omega' i/2))d\tau \Big),$$

with $\tau_l = i\omega'\epsilon/2 - \alpha'_l$, $\tau^*_l = i\omega'\epsilon/2 - \bar{\alpha}'_l$, where the integrals are taken along the lines parallel to the real axis of the τ plane. By integrating the function $\zeta(\tau)i\omega'/2 - \tau\zeta(i\omega'/2)$ on the parallelogram $(-\omega/2 + i\omega'/2, \tau_0, \tau_0 + \omega, \omega/2 + i\omega'/2)$, we obtain

$$\text{(6.21)} \quad \int_{\tau_0}^{\tau_0+\omega} (\zeta(\tau)\omega' i/2 - \tau\zeta(\omega' i/2))d\tau = \int_{(-\omega+i\omega')/2}^{(\omega+i\omega')/2} (\zeta(\tau)\omega' i/2 - \tau\zeta(\omega' i/2))d\tau - \pi i(\tau_0 + \omega/2 - i\omega'/2) - \epsilon'\pi\omega',$$

with $\epsilon' = 0$ for $\omega' > \Im\tau_0 > 0$, and $\epsilon' = 1$ for $-\omega' < \Im\tau_0 < 0$. Thus

$$\delta\phi = (1/\omega')\Re \sum_{l=1}^{h} [-\pi i(i\omega'\epsilon/2 - \bar{\alpha}'_l + \omega/2 - i\omega'/2) + \pi i(i\omega'\epsilon/2 - \alpha'_l + \omega/2 - i\omega'/2) - (1-\epsilon)\pi\omega'],$$

$$\text{(6.22)} \qquad \delta\phi = (-1/\omega') \sum_{l=1}^{h} 2\pi\Im\alpha'_l - h(1-\epsilon)\pi.$$

The interior boundary L_2, which has no point at infinity, cuspidal point, or point of inflection, will not cut itself if and only if $\delta\phi = +2\pi$, when the point $z = F(T)$ traces this curve, i. e. when $\epsilon = 1$. For this, it is obviously necessary and sufficient to have

$$\text{(6.23)} \qquad \sum_{l=1}^{h} \Im\alpha'_l = -\omega'.$$

Now, if this condition is satisfied, $\delta\phi$ will be equal to $2\pi - h\pi$ when the point $z = F(T)$ traces the exterior curve L_1, which has h infinite branches. From this, one can easily conclude that L_1 does not cut itself.

The functions $F'(T)$ which correspond to the solutions of our problem for the special case considered in this paragraph are therefore of the form

$$(6.24) \qquad F'(T) = A \prod_{l=1}^{h} (\sigma(T - \alpha_l)/\sigma(T - \beta_l))$$

where $0 \leq \beta_1 < \beta_2 < \cdots < \beta_h < \omega$; $\alpha_1 = \alpha'_1 + i\omega', \alpha_2 = \alpha'_2 \cdots, \alpha_h = \alpha'_h$, *with* $-\omega'/2 < \Im\alpha'_l < 0$, *and* $\sum_{l=1}^{h} \alpha_l - \sum_{l=1}^{h} \beta_l = 0$. *It follows from the preceding analysis that such a function satisfies the conditions* V, VI *of* § 4. *It satisfies obviously by construction also the conditions* I, II *and* IV *of that paragraph.*

Let us show now that the zeros of any function of the form

$$(6.25) \qquad F'(T) = (-2/\omega)\zeta(\omega/2) \sum_{l=1}^{h} f_l \exp(i\phi_l)\beta_l - \sum_{l=1}^{h} f_l \exp(i\phi_l)\zeta(T - \beta_l),$$

with $\sum_{l=1}^{h} f_l \exp(i\phi_l) = 0$, $0 \leq \beta_1 < \beta_2 < \cdots < \beta_h < \omega$, *and* (§ 4, III) $f_l > 0$, $0 \leq \phi_2 - \phi_1 \leq \phi_3 - \phi_1 \leq \cdots \leq \phi_h - \phi_1 < 2\pi$ *satisfy the conditions above, where such a function has h zeros* $\alpha'_1, \alpha'_2, \cdots, \alpha'_h$ *in* $-\omega'/2 < \Im T < 0$, $0 \leq \Re T < \omega$ *and* $\sum_{l=1}^{h} \Im\alpha'_l = -\omega'$. We consider first the function

$$(6.26) \qquad F'_0(T) = (\sigma(T + i\omega'/h - i\omega')/\sigma(T)) \prod_{l=1}^{h-1} (\sigma(T + i\omega'/h - l\omega/h)/\sigma(T - l\omega/h)),$$

$$= -\exp(-2\zeta(i\omega'/2)(T + i\omega'/h - i\omega'/2)) \prod_{l=0}^{h-1} (\sigma(T + i\omega'/h - l\omega/h)/\sigma(T - l\omega/h)),$$

which satisfies the equation

$$(6.27) \qquad F'_0(T + \omega/h) = F'_0(T)\exp(2\pi i/h).$$

The coefficient A_j of the principal part of this function at the pole $(j-1)\omega/h$ is obviously the ratio of

$$-\exp(-2\zeta(i\omega'/2)((j-1)\omega/h + i\omega'/h - i\omega'/2)) \prod_{l=0}^{h-1} \sigma((j-1)\omega/h + i\omega/h - l\omega/h)$$

to

$$\sigma((j-1)\omega/h)\sigma((j-2)\omega/h)\cdots\sigma(\omega/h)\sigma(-\omega/h)\cdots\sigma((j-h)\omega/h);$$

we have

$$A_j = A_1\exp(2(j-1)\pi i/h). \tag{6.28}$$

The function

$$F'_{0'}(T) = (2A_1/\omega)\sum_{j=1}^{h}\exp(2(j-1)\pi i/h) \times (\zeta(T-((j-1)\omega/h)\omega/2-(T-(j-1)\omega/h)\zeta(\omega/2)) \tag{6.29}$$

differs therefore from $F'_0(T)$ by a constant K. But it is easy to verify that the function $F'_{0'}(T)$ is also a solution of the linear functional equation (6.27). The constant K is therefore zero. The zeros $\alpha'_{10}=-i\omega'/h$, $\alpha'_{20}=-i\omega'/h+\omega/h,\cdots,\alpha'_{h0}=-i\omega'/h+(h-1)\omega/h$ of the function $F'_{0'}(T)=F'_0(T)$, which is of the form (6.25), are actually in the rectangle $-\omega'/2<\Im T<0$, $0\leq\Re T<\omega$, and we have $\sum_{j=1}^{h}\Im\alpha'_{j0}=-\omega'$.

We consider now an arbitrary function

$$F'(T) = (-2/\omega)\sum_{j=1}^{h} f_j\exp(i\phi_j)(\zeta(T-\beta_j)\omega/2-(T-\beta_j)\zeta(\omega/2))$$

of the form (6.25). This function has h zeros $\alpha'_1,\alpha'_2,\cdots,\alpha'_h$, which are continuous functions of $\phi_1,\phi_.,\cdots,\phi_h$; $f_1,f_2,\cdots,f_h$; $\beta_1,\beta_2,\cdots,\beta_h$, and which take respectively the values $\alpha'_{10},\alpha'_{20},\cdots,\alpha'_{h0}$ for

$$\phi_j=\phi_{j0}=2(j-1)\pi/h,\qquad f_j=f_{j0}=1,\qquad \beta_j=\beta_{j0}=(j-1)\omega/h \quad (j=1,2,\cdots,h).$$

From

$$\sum_{j=1}^{h}\Im\alpha'_j\equiv 0 \mod \omega',\qquad \sum_{j=1}^{h}\Im\alpha'_{j0}=-\omega',$$

we deduce by continuity

$$\sum_{j=1}^{h}\Im\alpha'_j=-\omega'.$$

We will show now that the points $\alpha'_1,\alpha'_2,\cdots,\alpha'_h$, which are initially inside of the strip $-\omega'/2<\Im T<0$, always remain in that strip. Suppose they are not inside of that strip for $\phi_j=\phi_{j1}$, $f_j=f_{j1}$, $\beta_j=\beta_{j1}$, and let us consider a system of continuous functions $\phi_j(\lambda),f_j(\lambda),\beta_j(\lambda)$ $(0\leq\lambda\leq 1)$ which satisfy the conditions

$$0 \leqq \phi_2(\lambda) - \phi_1(\lambda) \leqq \phi_3(\lambda) - \phi_1(\lambda) \leqq \cdots \leqq \phi_h(\lambda) - \phi_1(\lambda) < 2\pi,$$

$$f_j(\lambda) > 0, \qquad \sum_{j=1}^{h} f_j(\lambda) \exp(i\phi_j(\lambda)) = 0,$$

$$0 \leqq \beta_1(\lambda) < \beta_2(\lambda) < \cdots < \beta_h(\lambda) < \omega,$$

$$\phi_j(0) = \phi_{j0}, \qquad f_j(0) = f_{j0}, \qquad \beta_j(0) = \beta_{j0},$$

$$\phi_j(1) = \phi_{j1} \qquad f_j(1) = f_{j1}, \qquad \beta_j(1) = \beta_{j1},$$

at least one of the differences $\phi_j(\lambda) - \phi_1(\lambda)$ being different from 0 and π. The numbers $\alpha'_j = \alpha'_j(\lambda)$ are then continuous functions of λ which take the values $\alpha'_{10}, \alpha'_{20}, \cdots, \alpha'_{h0}$ for $\lambda = 0$. Some of the values of $\alpha'_1(\lambda), \alpha'_2(\lambda), \cdots, \alpha'_h(\lambda)$ for $\lambda = 1$ being outside of the strip $-\omega'/ < \Im T < 0$, we can find a value λ_0 $(0 < \lambda_0 \leqq 1)$ of the variable λ such that the points $\alpha'_1(\lambda)$, $\alpha'_2(\lambda), \cdots, \alpha'_h(\lambda)$ lie inside of that strip for $0 \leqq \lambda < \lambda_0$, and some of the points $\alpha'_1(\lambda_0), \alpha'_2(\lambda_0), \cdots, \alpha'_h(\lambda_0)$ become boundary points of $-\omega'/2 < \Im T < 0$, the others being still inside of it. Let us consider now the expressions

$$F(T, \lambda) = (-2/\omega) \int \sum_{j=1}^{h} f_j(\lambda) \exp(i\phi_j(\lambda)) (\zeta(T - \beta_j(\lambda))\omega/2 - (T - \beta_j(\lambda))\zeta(\omega/2)) dT,$$

$$F'(T, \lambda) = A \exp(-2\zeta(i\omega'/2)T) \prod_{j=1}^{h} \sigma(T - \alpha'_j(\lambda))/\sigma(T - \beta_j(\lambda)).$$

Since the points $\alpha'_j(\lambda)$ are all in $-\omega'/2 < \Im T < 0$ for $\lambda < \lambda_0$, we can say, according to what we have shown before, that $\arg F'(t + i\omega'/2, \lambda) = \phi$ changes monotonically from $\arg F'(t_0 + i\omega'/2, \lambda)$ to $\arg F'(t_0 + i\omega'/2, \lambda) + 2\pi$ if the real variable t varies from t_0 to $t_0 + \omega$, for $\lambda < \lambda_0$. By continuity we can say the same for $\lambda = \lambda_0$ with the only difference that if, for example, $\alpha'_1(\lambda_0)$ is on the line $\Im T = -\omega'/2$, then the expression $\arg F'(t + i\omega'/2, \lambda_0)$ makes a jump of magnitude π at the point $\alpha'_1(\lambda_0) + i\omega'$ of the segment $(t_0 + i\omega'/2, t_0 + i\omega'/2 + \omega)$. On the other hand, the point $F(t + i\omega'/2, \lambda_0)$ traces a closed curve when t varies from t_0 to $t_0 + \omega$. But this is only possible if this closed curve is a line segment traced twice and that would be only possible if all of the numbers $\phi_2(\lambda_0) - \phi_1(\lambda_0), \phi_3(\lambda_0) - \phi_1(\lambda_0), \cdots, \phi_h(\lambda_0) - \phi_1(\lambda_0)$ were equal to zero or π. The points $\alpha'_1(\lambda_0), \alpha'_2(\lambda_0), \cdots, \alpha'_h(\lambda_0)$ are therefore all in the strip $-\omega'/2 < \Im T \leqq 0$. It remains to show that none of them can be on the line $\Im T = 0$.

Since the points $\alpha'_j(\lambda)$ are all in the strip $-\omega'/2 < \Im T < 0$ for $\lambda < \lambda_0$, $\arg F'(t, \lambda)$ changes monotonically from $\arg F'(t_0, \lambda)$ to $\arg F'(t_0, \lambda) + 2\pi - h\pi$,

when the real variable t changes from t_0 to $t_0 + \omega$; also it changes from $\phi_l - k\pi$ to $\phi_{l+1} - (k+1)\pi$ when t traces the interval between two consecutive poles $\beta_l(\lambda) + N\omega$ and $\beta_{l+1} + M\omega$ of $F'(T, \lambda)$. By continuity one can say the same for $\lambda = \lambda_0$ with the only difference that if, for example, $\alpha_1(\lambda_0)$ is in the interval $(\beta_l(\lambda_0) + N\omega, \beta_{l+1}(\lambda_0) + M\omega)$, then the expression $\arg F'(t, \lambda_0)$ will make a jump of magnitude $-\pi$ at the point $\alpha_1(\lambda_0)$. We shall have therefore

$$(\phi_{l+1}(\lambda_0) - (k+1)\pi) - (\phi_l(\lambda_0) - k\pi) < -\pi,$$

or $\phi_{l+1}(\lambda_0) - \phi_l(\lambda_0) < 0$, which contradicts our hypothesis $\phi_{l+1}(\lambda_0) > \phi_l(\lambda_0)$.

We can now summarize the results of this paragraph as follows.

One obtains a solution of our problem, for $m = 2$ and σ_x, σ_y, τ finite everywhere in C, by putting:

$$z = F(T) = (-2/\omega) \sum_{j=1}^{h} f_j \exp(i\phi_j) \int (\zeta(T - \beta_j)\omega/2 - (T - \beta_j)\zeta(\omega/2))dT,$$

$$= - \sum_{j=1}^{h} f_j \exp(i\phi_j) \log \theta_1((T - \beta_j)/\omega),$$

with $0 \leq \phi_2 - \phi_1 \leq \phi_3 - \phi_1 \leq \cdots \leq \phi_h - \phi_1 < 2\pi$, $0 \leq \beta_1 < \beta_2 < \beta_3 < \cdots \leq \beta_h < \omega$, $f_j > 0$, $\sum_{j=1}^{h} f_j \exp(i\phi_j) = 0$; *and* $H = \bar{F}'(T)/F'(T)$

$$= \sum_{j=1}^{h} f_j \exp(-i\phi_j)(\zeta(T - \beta_j)\omega/2 - (T - \beta_j)\zeta(\omega/2))$$

$$\times \sum_{j=1}^{h} f_j \exp(i\phi_j)(\zeta(T - \beta_j)\omega/2 - (T - \beta_j)\zeta(\omega/2))$$

$$\sigma_x = 2\alpha(1 + \Re H(T)), \qquad \sigma_y = 2\alpha(1 - \Re H(T)), \qquad \tau = -2\alpha \Im H(T).$$

The domain C is then the image of the rectangle $\omega'/2 > \Im T > 0$, $0 \leq \Re T < \omega$, in the z plane, and every solution of that problem can be obtained in this way.

Appendix.

Let ω, ω' be real and positive; then the following inequalities hold for the Weierstrass ζ function corresponding to the periods ω and $i\omega'$:

$$\Re(\zeta(T)i\omega'/2 - T\zeta(i\omega'/2)) \geq 0 \quad \text{for } n\omega' \leq \Im T \leq \omega'/2 + n\omega',$$

$$\Re(\zeta(T)i\omega'/2 - T\zeta(i\omega'/2)) \leq 0 \quad \text{for } -\omega'/2 + n\omega' \leq \Im T \leq n\omega',$$

and equality holds only on the boundaries of the strips.

Proof. Because of the relations

$$
\begin{aligned}
\zeta(T+\omega)i\omega'/2 - (T+\omega)\zeta(i\omega'/2) &= \zeta(T)i\omega'/2 - T\zeta(i\omega'/2) + i\omega'\zeta(\omega/2) - \omega\zeta(i\omega'/2),\\
&= \zeta(T)i\omega'/2 - T\zeta(i\omega'/2) + \pi i,\\
\zeta(T+i\omega')i\omega'/2 - (T+i\omega')\zeta(i\omega'/2) &= \zeta(T)i\omega'/2 - T\zeta(i\omega'/2) + i\omega'\zeta(i\omega'/2) - i\omega'\zeta(i\omega'/2)\\
&= \zeta(T)i\omega'/2 - T\zeta(i\omega'/2),
\end{aligned}
$$

it is sufficient to prove the inequalities for $n = 0$ and $-\omega/2 < \Re T \leq \omega/2$. According to the relation of Legendre, the function

$$g(T) = \zeta(T)i\omega'/2 - T\zeta(i\omega'/2)$$

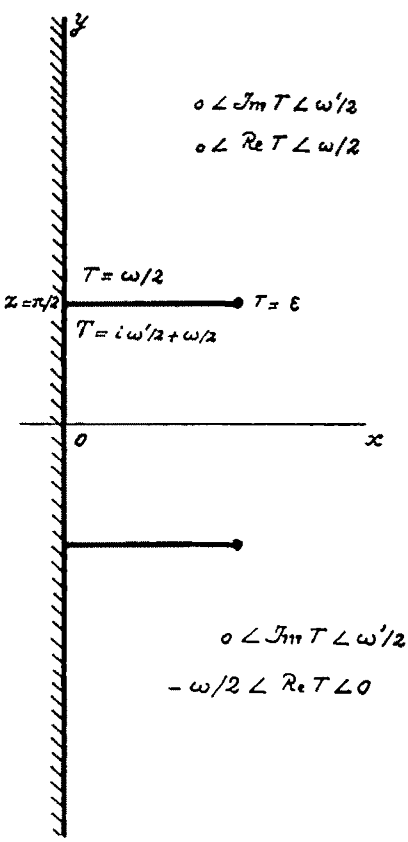


FIG. 4.

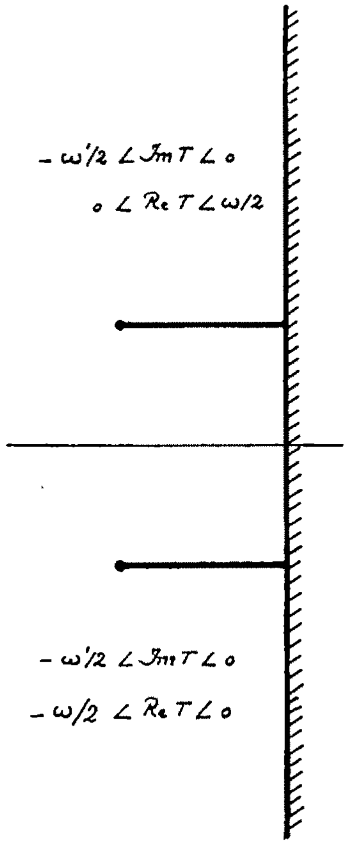


FIG. 5.

takes the same value $i\pi/2$ for $T = \omega/2$ and for $T = (\omega + i\omega')/2$. Its derivative

$$g'(T) = -(\wp(T)i\omega'/2 + \zeta(i\omega'/2))$$

is purely imaginary on the segment $(\omega/2, (\omega + i\omega')/2)$. This derivative has therefore a zero $T = \epsilon$ on this segment. The other zero of $g'(T)$ in the period rectangle $(-i\omega'/2, -i\omega'/2 + \omega, i\omega'/2 + \omega, i\omega'/2)$ is then $T = \bar{\epsilon}$. The function $z = g(T)$ therefore maps the rectangle $(0, \omega/2, (\omega + i\omega')/2, i\omega'/2)$ conformally on the domain shown in fig. 4. By symmetry with respect to the segment $(i\omega'/2,\ 0)$ we see that the rectangle $(-\omega/2,\ \omega/2,\ (\omega + i\omega')/2,\ -\omega/2 + i\omega'/2)$ is mapped on the right side of the imaginary axis of the z plane, while the rectangle $(-\omega/2,\ \omega/2,\ \omega/2 - i\omega'/2,\ -\omega/2 - i\omega'/2)$ is mapped on the left side of this axis (symmetrically with respect to the segment $(-\omega/2, \omega/2)$), and our inequalities then become obvious.

UNIVERSITY OF ISTANBUL AND
UNIVERSITY OF MARYLAND,
INSTITUTE FOR FLUID DYNAMICS AND APPLIED MATHEMATICS.

A PERIODIC ALGORITHM FOR CUBIC FORMS.*

By Harvey Cohn.

1. Introduction. In a very famous memoir in 1896, Minkowski [5] set up a continued fraction type of algorithm for the product of three real linear homogeneous forms in three integral variables. By the precedent set in the ordinary theory of binary quadratic forms, such an algorithm has to supply a definition of *reduced form* with several properties: First of all every form should be equivalent under unimodular change of variables to one or more reduced forms. Furthermore a concept of *neighboring form* should be present so that the reduced forms will arrange themselves into chains. Finally when the three linear forms are conjugates, each spanning a (totally real) *cubic module*, the number of distinct reduced forms should be *finite*. Probably the most immediate application of such an algorithm would be the determination of units of a totally real cubic field as automorphs within the finite chain of forms.

The purpose of this work is to present a new algorithm for such cubic forms, combining some of the features of an ordinary (modified) continued fraction algorithm with some of the features of Minkowski's algorithm. This new algorithm follows the former one to the extent that reduction is based on a visually convenient array of $+$ and $-$ signs rather than on "admissible" bodies; and it follows the latter algorithm to the extent that the proof given here for the reduction theorem uses Minkowski's work as a convenient starting point. Once the theoretical aspects are established, it will be seen that the actual calculations in cubic fields are very simple, although the units obtained there are far from thoroughly investigated.

2. Modified simple continued fraction algorithm. The familiar simple continued fraction (for binary quadratic forms) of course fulfils the above requirements *par excellence*. We present a modification to emphasize some curious resemblances to the definition of reduced forms that we presently give for cubics. The quadratic form may be written

$$\Phi(n, m) = x_1 x_2, \tag{2}$$

* Received August 27, 1951. Research currently sponsored by the U. S. Army O. O. R.

where $x_i = n - \theta_i m$, $(n, m$ integers$)$, $i = 1, 2$ and θ_i are conjugate real quadratics. The condition for reduction (as stated by Gauss) is that θ_1 and $-1/\theta_2$ be each greater than 1. If we consider the lattice in the (x_1, x_2)-plane we find this is *almost the condition that* V_1, V_2, V_3, *the* (x_1, x_2)*-vectors corresponding to* $(n, m) = (0, 1)$, $(1, 0)$, $(-1, -1)$ *respectively, do not all three lie in one pair of opposite quadrants of the* (x_1, x_2)*-plane*. Here note that V_1, V_2, V_3 have the property that any two of the three are basis vectors of the lattice, and $V_1 + V_2 + V_3 = 0$. The idea of using not two but *three* such vectors (but in another formulation), goes back to Selling [1]. The concept of neighbor in the simple continued fraction is (again) *almost the condition that two of the three vectors* V_1, V_2, V_3 *be common to the two forms in question*. The resemblance of this modified continued fraction algorithm to the one presented in this paper will become clear as matters develop.

3. Minkowskian basis. We now use as a point of departure the existence of a Minkowskian basis, an immediate consequence of Minkowski's algorithm.

We recall that a lattice is said to be *admissible* with respect to a convex body K (or vice versa) if no point of the lattice except the origin is interior to K. The convex body which we use is the rectangular parallelopiped:

$$P(\xi_1, \xi_2, \xi_3) : \; |x_i| \leq \xi_i. \tag{3.1}$$

The lattice we use can be written as

$$x_i = b_{i1}m_1 + b_{i2}m_2 + b_{i3}m_3, \qquad m_j \text{ integral}, i, j = 1, 2, 3. \tag{3.2}$$

Its determinant D is assumed $\neq 0$, and we henceforth make the simplifying assumption that *none of the components* x_1, x_2, *or* x_3 *shall vanish for any vector of the lattice except the origin.* This fits in conveniently with the intended application, where the triple (b_{i1}, b_{i2}, b_{i3}) spans the totally real cubic field K_i (where the index i denotes the conjugate).

We consider an ordered triple of lattice vectors (A_1, A_2, A_3), where $A_i = (x_1^{(i)}, x_2^{(i)}, x_3^{(i)})$. We call such a triple *Minkowskian* if $P(x_1^{(1)}, x_2^{(2)}, x_3^{(3)})$ is admissible. Such triples exist under very broad assumptions ([2]) and *from Minkowski's algorithm* ([5]) *such a triple is either co-planar or a lattice basis.* (The proof consists essentially, [4], of eliminating the possibility that the triple generate a sub-lattice of index 2, and this can be done more simply by the methods of **5** below).

Minkowski's algorithm also supplies a process of forming neighbors which we need not discuss in detail except for the following points of imme-

diate concern: The process [3] of "lowering the faces" shows that every Minkowskian triple has exactly *three* neighbors. The process, in addition, involves a "visualization" of the lattice with a view to admissibility. (The neighbor process proposed in **10** below will be less simple in the first respect but much more simple in the second.)

Minkowski's neighbor process is also such that a co-planar Minkowskian triple (fortunately) cannot have co-planar neighbors ([3]); hence we have the

THEOREM. *Every lattice has a Minkowskian triple that serves as a basis.*

We call this a *Minkowskian basis* and proceed from there.

Reduction Theorem.

4. Definition of reduced form. We start off by considering a 3×4 *matrix* ϕ written as four column vectors P, Q, R, S with three x_i components p_i, q_i, r_i, s_i:

$$\tag{4.1} \phi: \begin{array}{c c} & \begin{array}{cccc} P & Q & R & S \end{array} \\ \begin{array}{c} (x_1:) \\ (x_2:) \\ (x_3:) \end{array} & \left\{\begin{array}{cccc} p_1 & q_1 & r_1 & s_1 \\ p_2 & q_2 & r_2 & s_2 \\ p_3 & q_3 & r_3 & s_3 \end{array}\right\} \end{array}.$$

Such a matrix is said to be reduced if

$$\tag{4.21} O = P + Q + R + S$$

and if

$$\tag{4.22} \pm \operatorname{sgn} \phi = \left\{\begin{array}{cccc} + & - & - & + \\ - & + & - & + \\ - & - & + & + \end{array}\right\};$$

(i. e., if the components of ϕ have either the exact or reverse pattern of $+$ signs and $-$ signs of the last matrix). Such a matrix is written out in **11** (below).

From the definition it is clear that no three of the column vectors P, Q, R, S are co-planar (for, if so, there would be a common perpendicular to all four of them in contradiction to the fact that these four vectors and their negatives lies one in each octant).

We next define classes of *proportional* matrices, by saying ϕ and ϕ' are proportional, i. e.

$$\tag{4.31} \phi \sim \phi',$$

if the three (x_i) rows of each are proportional, $p_i/p'_i = q_i/q'_i = r_i/r'_i = s_i/s'_i = w_i (\neq 0)$. We say that ϕ and ϕ' are *strictly proportional*, or

$$\phi \approx \phi', \tag{4.32}$$

if $\phi \sim \phi'$ and, in addition, $w_1 w_2 w_3 = 1$.

Clearly if $\phi \sim \phi'$ and ϕ' is reduced, then ϕ^λ is reduced, where λ is a permutation in the Klein four-group on (P, Q, R, S). (If all constants, w_i, are positive, then λ is the identity.)

We next consider *degenerate ternary forms*, i. e.

$$\Phi = \sum_{ikl}^{3} = a_{ikl} m_i m_k m_l, \tag{4.41}$$

where

(4.42)

$$\Phi = x_1 x_2 x_3; \qquad x_i = b_{i1} m_1 + b_{i2} m_2 + b_{i3} m_3, \qquad 1 \leq i, j \leq 3;$$
$$B_j = (b_{1j}, b_{2j}, b_{3j}).$$

We call such a form *reduced*, or we say the three column vectors B_1, B_2, B_3 *are a reduced basis*, if these three vectors, taken in some order, come from the column vectors P, Q, R, S of a reduced 3×4 matrix. This definition is immediately invariant under permutation of variables m_1, m_2, m_3. This definition is such that given a form Φ, the question of whether or not it is reducible does not depend on how the factors x_i are identified. In fact if the factors x_i are transformed by the appropriate group: $x_1 \to c_1 x_i$, $x_2 \to c_2 x_j$; $x_3 \to c_3 x_k$, $i \neq j \neq k$, $(c_1 c_2 c_3 = 1)$, the corresponding 3×4 matrix obeys $\phi \to \phi'$, where

$$\phi' \approx \phi^\sigma, \tag{4.5}$$

and σ denotes a permutation from the alternating group on (P, Q, R, S).

We could finally discuss a *degenerate quaternary form* given by formulas similar to (4.41) and (4.42) except that the index j would go from 1 to 4 and there also would be 4 values of m_j. Then a *reduced* form would be one in which the B_1, B_2, B_3, B_4 were, in some order, the P, Q, R, S of a reduced 3×4 matrix. Thus any reduced degenerate *ternary* form Φ results from a reduced degenerate *quaternary* form in which the vanishing of a single m_i leaves the form equal to Φ (in the three remaining variables). The quaternary forms are such that each *strict* proportionality class of ϕ determines the same form (to within permutation of the m_i) together with exactly all the ϕ^σ described in (4.5). The table in **13** (below) will provide instances where

4 or 12 such ϕ determine the same quaternary form. Such properties necessarily extend to the ternary forms.

5. Proof that every degenerate ternary form is equivalent to a reduced form.

THEOREM. *If A_1, A_2, A_3 are a Minkowskian basis, then $\pm A_1$, $\pm A_2$, $\pm A_3$, with an appropriate choice of signs, is a reduced basis.*

Remark. It seems that some significance was attached to the 3×3 array of signs of $\pm A_i$ by Minkowski [5], but the role of a fourth vector was not considered.

Proof. We can without loss of generality multiply the A_i by constants and reflect the coordinate planes so as to have

$$(5.11) \qquad a_{ii} = 1, \quad |a_{ij}| < 1 \qquad \text{if } i \neq j, a_{21} > 0, a_{31} > 0,$$

$$(5.12) \qquad A_1 = (1, a_{12}, a_{13}), \quad A_2 = (a_{21}, 1, a_{23}), \quad A_3 = (a_{31}, a_{32}, 1).$$

Now the hypothesis that A_1, A_2, A_3 is a Minkowskian basis requires, among other things, that the 20 vectors $\pm A_i \pm A_j$ $(i \neq j)$, $\pm A_1 \pm A_2 \pm A_3$ lie on or outside the cube $|x_i| < 1$. So from here on, we *weaken the hypothesis,* (see [4]) so that the a_{ij} satisfy the conditions in (5.11) together with merely

$$(5.2) \qquad \begin{aligned} &\max_k |\pm a_{ik} \pm a_{jk}| > 1 \qquad (i \neq j), \\ &\max_k |\pm a_{1k} \pm a_{2k} \pm a_{3k}| > 1, \text{ for each choice of } \pm. \end{aligned}$$

We now assert that the pattern of signs of the four vectors

$$(5.3) \qquad \pm A_1, \pm A_2, \pm A_3, \mp A_1 \mp A_2 \mp A_3,$$

does in some order accord with the system (4.22), if the $\pm$ signs are chosen properly for each $\pm A_i$.

Next we visualize the $\Pi_1: (a_{21}, a_{31})$-plane, the $\Pi_2: (a_{12}, a_{32})$-plane, and the $\Pi_3: (a_{13}, a_{23})$-plane, written more generally as (ξ, η)-planes. We see that the regions determined by the inequalities (5.11) and (5.2) are bounded by lines of the types: $\pm \xi = 0, 1$; $\pm \eta = 0, 1$; $\pm \xi \pm \eta = 0, 1$. Thus in the $\Pi_1 \times \Pi_2 \times \Pi_3$ space, $4 \times 16 \times 16 = 1024$ regions with isosceles right-triangular projections are determined. In fact, considering the sixteen Π_2- and the sixteen Π_3-projection triangles, we see that each one contains in its interior exactly one point of one of the 16 types $(\pm \frac{1}{2}, \pm \frac{1}{4})$, $(\pm \frac{1}{2}, \pm \frac{3}{4})$, $(\pm \frac{1}{4}, \pm \frac{1}{2})$, $(\pm \frac{3}{4}, \pm \frac{1}{2})$, while the Π_1-projection triangle contains only of the four types

formed using only $+$ signs, $(a_{k1} > 0)$. In each of these 1024 (open) regions the inequalities in (5.11) are always true, while the inequalities in (5.2) are *each* always true or always false. In the 184 cases where (5.11) and (5.2) are *always* true it is easy to choose the vectors in (5.3) to have the proper array of signs. The handling of all 1024 cases can be easily accomplished in a few hours by using $4+16+16$ strips of paper with the x_1-, x_2-, and x_3-components of the 26 vectors $\pm A_i$, $\pm A_i \pm A_j$, $\pm A_1 \pm A_2 \pm A_3$, $(i \neq j)$.

6. Alternative form of the reduction theorem. We have shown that every basis is equivalent under change of basis to a reduced basis. It is no loss of generality to assume that the change of basis involves a matrix of determinant $+1$ (as *all* basis vectors can be multiplied by -1 without affecting the definition of reduction). Likewise we may assume by (4.21) that B_1, B_2, B_3 are S, P, Q respectively. Finally we divide each x_i by a constant so as to put the lattice in inhomogeneous form, whence $b_{1i} = 1$. Then calling $b_{2i} = \xi_i$, $b_{3i} = \eta_i$ we obtain the following result:

Let three non-collinear points (ξ_i, η_i) of the projective $(\xi\eta)$-plane be given in some fixed order. It is possible to find an integral unimodular transformation such that the points (ξ_i, η_i) simultaneously go into the points (ξ'_i, η'_i),

$$(6.1)\qquad \xi'_i = \frac{a_1 + b_1\xi_i + c_1\eta_i}{a_3 + b_3\xi_i + c_3\eta_i}, \quad \eta'_i = \frac{a_2 + b_2\xi_i + c_2\eta_i}{a_3 + b_3\xi_i + c_3\eta_i}, \quad \det(a_1 b_2 c_3) = +1$$

lying in the following three regions of the projective plane respectively:

$$(6.2)\qquad \begin{aligned} R_1:&\ \ \xi_1 \geqq 0, \eta_1 \leqq 0, & 1 + \xi_1 + \eta_1 \geqq 0;\\ R_2:&\ \ \xi_2 \leqq 0, \eta_2 \geqq 0, & 1 + \xi_2 + \eta_2 \geqq 0;\\ R_3:&\ \ \xi_3 \leqq 0, \eta_3 \leqq 0, & 1 + \xi_3 + \eta_3 \leqq 0. \end{aligned}$$

Note that in order to cope with our previous lattice restriction (that no lattice vector have a zero component except the origin), we must extend the regions to the projective plane, thus permitting division by zero.

7. Cyclic basis. A very important type of lattice is one with a cyclic basis,

$$(7.1)\qquad b_{ij} = \alpha_k, \text{ where } k \equiv i + j \pmod 3 \qquad (k = 1, 2, 3).$$

In this case the ternary form Φ, as a function of the m_i, is invariant under cyclic permutations, and conversely. Here the basis B_j, $(j = 1, 2, 3)$, (in the notation of **4**) is reduced if and only if, of the four quantities α_1, α_2, α_3, $-(\alpha_1 + \alpha_2 + \alpha_3)$, two are positive and two are negative,

$$(\text{or } \alpha_1\alpha_2\alpha_3(\alpha_1 + \alpha_2 + \alpha_3) < 0).$$

We now single out the "hard case," where, under a change of indices and (possibly) a change in sign of all α_i, we have the inequalities

$$(7.2) \qquad \alpha_1 + \alpha_2 > \alpha_3 > \alpha_2 > \alpha_1 > 0.$$

Then, except for the hard case just cited, the cyclic basis will be finitely close to a reduced basis in the sense that the vectors P, Q, R, S for the reduced basis can be chosen from among the 26 vectors $\pm B_i$, $\pm B_i \pm B_j$, $\pm B_1 \pm B_2 \pm B_3$, $(i \neq j)$.

The above statement can be easily verified. We further note that the hard case must be separated since the three numbers α_1, α_2, α_3 can otherwise be arbitrarily close to one another and still satisfy the conditions (7.2). Therefore, even if the set of 26 vectors above were extended to any finite set, however large, of the type $k_jB_1 + l_jB_2 + m_jB_3$, it would still be possible to choose the α_i so that of the four desired basis vectors P, Q, R, S, the one (S) which is in the all $+$ (or all $-$) octant would be the one for which $k_j + l_j + m_j \neq 0$, while the other three vectors would be such that $k_j + l_j + m_j = 0$, contradicting condition (4.21).

Fortunately, in many applications, the hard case simply does not occur. For instance, for later reference, take

$$(7.3) \qquad \alpha_k = \sum \exp(2\pi i m/p), \qquad (k = 1, 2, 3),$$

where p $(\equiv 1 \bmod 3)$ is a prime and m through a complete set of residues of cubic character indicated by k. Here $\alpha_1 + \alpha_2 + \alpha_3 = -1$, and therefore if the α_i were all of one sign (as in (7.2)), we should have $|\alpha_1\alpha_2\alpha_3| \leqq 1/27$ in contradiction to the fact that α is an algebraic integer. Hence the lattice given by (7.1) and (7.3), (determined by contraction of a cyclotomic field), has a reduced basis very close to the cyclic basis described here.

Finiteness.

8. Norms and discriminants. We now consider the case where the x_1, x_2, x_3 components in (4.42) are the conjugates of three numbers spanning a (totally real) cubic field. We shall show that although the neighbor process, for instance, produces a chain of 3×4 matrices (see **12** below), they are strictly proportional to only a finite number of classes. Thus the ternary and quaternary forms take on only a finite number of values.

As a preliminary step, multiply through the matrices so as to make all of the p_i, q_i, r_i, s_i algebraic integers. We let Δ represent the (common) absolute value of the determinants formed by omitting any one of the four

columns. Here Δ^2 is an integer which represents the field discriminant multiplied by j^2, where j represents the index of the sub-lattice (or module) of integers in the ring of all integers of the field. Of course Δ is invariant under change of basis.

In the usual notation, where sub-scripts denote conjugates, we denote the norm and discriminant of the vector $A = (a_1, a_2, a_3)$ by means of

$$(8.1)\quad N(A) = a_1a_2a_3; \qquad (8.2)\quad d(A) = [(a_1 - a_2)(a_2 - a_3)(a_3 - a\ )]^2.$$

Clearly $d(A) = 0$ only if $a_1 = a_2 = a_3 =$ rational number, while $N(A) = 0$ only if $a_1 = a_2 = a_3 = 0$. Furthermore when A is an algebraic integer, N and d are natural integers.

We shall next show that when Δ is bounded (and > 0) the norms of at least *three* of the four algebraic integers P, Q, R, S are bounded $(< (\Delta/3)^3)$ while the discriminants of *all twelve* fractions of the type $P/Q = (p_i/q_i)$ are likewise restricted to a finite set of rational fractions (whose numerators go from 1 to $[(\Delta/3)^6]$ and whose denominators go from 1 to $[(\Delta/3)^{18}]$).

The above conclusions follow from the expansion of Δ in (4.1) along (say) columns R and S using, each time, the minors of P and Q.

$$(8.3)\quad \begin{aligned} -\Delta &= (p_1q_1 - q_1p_2)r_3 + (p_3q_1 - q_3p_1)r_2 + (p_2q_3 - q_2p_3)r_1,\\ +\Delta &= (p_1q_2 - q_1p_2)s_3 + (p_3q_1 - q_3p_1)s_2 + (p_2q_3 - q_2p_3)s_1. \end{aligned}$$

Now take either the upper or the lower lines of (8.3) according as $(p_1q_2 - q_1p_2)$ is negative or positive. (Note that $p_iq_j - p_jq_i \neq 0$ for $i \neq j$ by the rational independence of P and Q.) In (say) the first case, reference to system (4.22) reveals that $-\Delta$ is the sum of three negative quantities, abbreviated $t_1 + t_2 + t_3$, whence easily,

$$(8.4)\qquad (\Delta/3)^3 \geqq |\, t_1t_2t_3 \,| = |\, N(R) \,| \cdot |\, G(P, Q) \,|,$$

where $G^2(P, Q) = (p_1q_2 - q_1p_2)^2(p_3q_1 - q_3p_1)^2(p_2q_3 - q_2p_3)^2$. Since $G^2(P, Q)$ is a rational integer, it follows that one column, R, of the chosen pair R, S, has bounded norm. By repeating the argument we see that some three of the column vectors (say) P, Q, R have bounded norm (as indicated above). Likewise, regardless of whether R or S has the bounded norm,

$$(8.5)\qquad G^2(P, Q) = N^4(Q)d(P/Q) = N^4(P)d(Q/P);$$

hence any of the twelve fractions of the type P/Q has, from (8.4), a discriminant restricted to the set of values indicated above.

9. Completion of finiteness proof. Using the previous information let us suppose that fixed values of $N(P/R)$, $N(Q/R)$, $d(P/R)$, $d(Q/R)$, $d(P/Q)$,

$d(S/R)$ are given as rational numbers $g_1, \cdots, g_6$ respectively, chosen from a finite set. We next see that then P/R and Q/R have only a finite set of values.

To see this we write $P/R = (u_i)$, $Q/R = (v_i)$, $P/Q = (u_i/v_i)$, $S/R = (-1 - u_i - v_i)$ by virtue of (4.21). Thus we find that

$$
(9.1)\quad
\begin{aligned}
&u_1u_2u_3 = g_1, \ldots \qquad v_1v_2v_3 = g_2,\\
&(u_1 - u_2)(u_2 - u_3)(u_3 - u_1) = \pm\sqrt{g_3},\\
&(v_1 - v_2)(v_2 - v_3)(v_3 - v_1) = \pm\sqrt{g_4},\\
&(u_1/v_1 - u_2/v_2)(u_2/v_2 - u_3/v_3)(u_3/v_3 - u_1/v_1) = \pm\sqrt{g_5},\\
&(u_1 + v_1 - u_2 - v_2)(u_2 + v_2 - u_3 - v_3)(u_3 + v_3 - u_1 - v_1) = \pm\sqrt{g_6}.
\end{aligned}
$$

This constitutes six equations in the six variables u_i, v_i. There are either no more than $3^5 \cdot 6 \cdot 2^4$ solutions or else infinitely many. The latter possibility is somewhat tedious to exclude, but if there were infinitely many solutions for some g_i (all $\neq 0$), it would be possible to parametrize a one dimensional manifold in terms of $t = u_1/u_3$ as (say) $t \to \infty$. By trying to calculate from (9.1) the order of magnitudes in t of the u_i and v_i, we see that this possibility leads to a contradiction.

Now knowing that the values of

$$p_i/r_i,\ q_i/r_i \text{ and } s_i/r_i \qquad (= -1 - p_i/r_i - q_i/r_i)$$

are finite in number we can complete the proof. For then the 3×4 matrices ϕ belong to only a finite number of proportionality classes. But since the number of values of $p_1p_2p_3$ $(= N(P))$ is limited, there must also be a finite number of strict proportionality classes. Finally (see **4**) we have a bound on the number of forms Φ.

Neighbors and Chains.

10. Definition of neighbors. Our definition of neighbors will be given as a definition of neighboring *reduced* 3×4 matrices ϕ, but it can be carried over to the corresponding ternary or quaternary forms Φ. Since the order of the column vectors P, Q, R, S is fixed by the array of signs, we need only supply the "parabolic" generator of the chain. In fact we take the simplest ones consistent with condition (4.21), namely those of the type

$$(10.1)\qquad \phi \to \phi' \quad (P \to P,\quad Q \to Q + P,\quad R \to R,\quad S \to S - P)$$

which are written for convenience as

$$(10.2) \qquad \phi(Q.P.S)\phi' \quad \text{or} \quad \phi'(S.P.Q)\phi.$$

Of the 24 matrices ϕ' thus formed, we call the *reduced* matrices *neighbors* of ϕ.

This process creates 3, 4, 5, or 6 neighbors, depending on the relative magnitude of the components in the matrix ϕ. Furthermore, (positively) proportional reduced matrices have the same neighbors. To calculate the neighbors is very easy in that it requires not even a mental picture of the lattice as did the Minkowskian algorithm. The automatic process will now be given in terms of an illustration.

11. Illustrative calculation. For instance consider the lattice with reduced basis, (compare 7), $b_{ij} = 2\cos 2\pi ij/7$. We obtain the array of (approximate) numerical values:

$$\begin{array}{lll} & & \begin{array}{cccc} P & Q & R & S \end{array} \\ \phi_0: & \begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array} & \left\{ \begin{array}{cccc} 1.25 & -\ .45 & -1.80 & 1 \\ -1.80 & 1.25 & -\ .45 & 1 \\ -\ .45 & -1.80 & -1.25 & 1\ . \end{array} \right. \end{array}$$

Looking at the x_1 row we find the elements of *like* sign compare as follows in numerical magnitude:

$$(11.1) \qquad |r_1| > |q_1|, \qquad |p_1| > |s_1|,$$

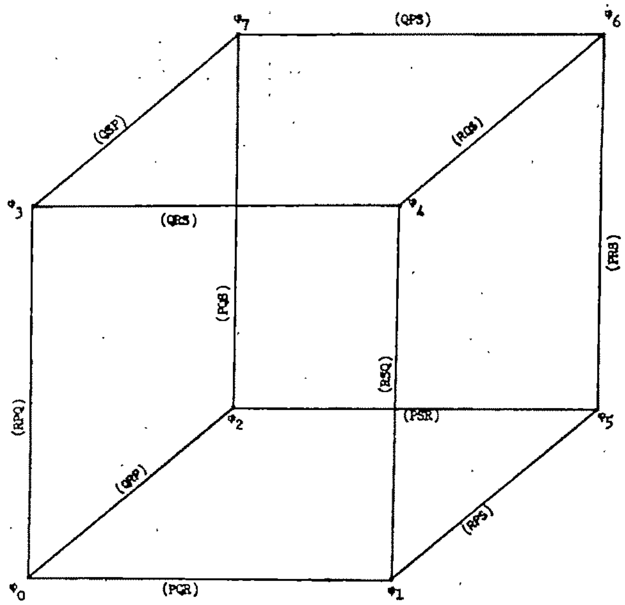
written as $(\ .R.Q)$, $(\ .P.S)$. The elements of unlike sign compare as follows in numerical magnitude:

$$(11.2) \qquad |p_1| < |r_1|, \quad |q_1| < |p_1|, \quad |q_1| < |s_1|, \quad |s_1| < |r_1|,$$

written as $(P.R.\ )$, $(Q.P.\ )$, $(Q.S.\ )$, $(S.R.\ )$ respectively.

Looking at the x_2 row we obtain six more entries, $(\ .P.R.)$, $(\ .Q.S.)$; $(Q.P.\ )$, $(S.P.\ )$, $(R.Q.\ )$, $(R.S.\ )$; and looking at the x_3 row we obtain six final entries, $(\ .Q.P.)$, $(\ .R.S)$; $(P.R.\ )$, $(P.S.\ )$, $(R.Q.\ )$, $(S.Q.\ )$.

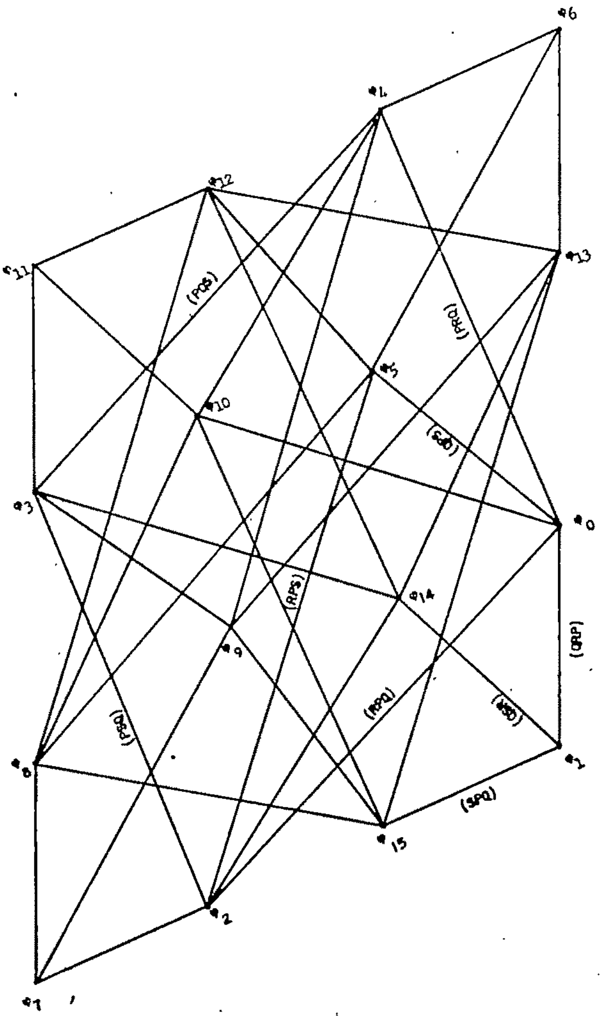
We now have 18 entries (possibly repetitive), and the rule is simply that the neighbors are those remaining when the triples of the above 18 types are *eliminated.* Thus going back (11.1) we see the entry $(\ .R.Q)$ *disposes of* $(P.R.Q)$ and $(S.R.Q)$ as neighbors, while in (11.2), the entry $(P.R.\ )$ *disposes* of $(P.R.Q)$ and $(P.R.S)$, etc. It is seen that only three neighbors for ϕ_0 are left, namely $\phi_0(P.Q.R)\phi_1$, $\phi_0(Q.R.P)\phi_2$, $\phi_0(R.P.Q)\phi_3$, which occur in the lower left hand corner of figure 1 below.



12. Closed chains. By use of this neighbor process we obtain a succession of 3×4 matrices ϕ. They must necessarily correspond to a finite set of strict proportionality classes. For instance, referring to the case we obtain the succession of neighbors $\phi_0(PQR)\phi_1(RPS)\phi_5(RSP)\phi_2(PRQ)\phi_8$; it is not hard to guess even from the approximations used above, that $\phi_8 \approx \phi_0$, and indeed it can be seen to check with the defining equations.

In this way we obtain by succession sub-sets of the matrices in each strict proportionality class (in fact, we obtain sub-sets with *positive* relative proportionality constants). Certain of these strict proportionality classes are neighbors to one another. Hence we may always draw these classes as points connected by lines indicating neighbors. This is done in figures 1 and 2 (q. v., also see tables below).

13. Further calculations. We submit now, five calculations of closed chains, summarized in the accompanying table. The modules taken are some of the integral modules with lowest discriminant. The case $\Delta^2 = 196$ is the sub-module of index 2 of $\Delta^2 = 49$, while $\Delta^2 = 148$ is, of course, non-abelian. The (x_i) conjugates are taken in an order consistent with the sign matrix (4.22).



Δ^2	Defining Equation	P	Q	R	S	Number of prop. classes
49	$\rho^3+\rho^2-2\rho-1=0$	ρ	$1-\rho-\rho^2$	$-2+\rho^2$	1	$2\times 4 \quad =8$
81	$\rho^3 \quad -3\rho-1=0$	ρ	$-2-\rho+\rho^2$	$1-\rho^2$	1	$1\times 4+1\times 12=16$
148	$\rho^3 \quad -4\rho+2=0$	$-3+\rho+\rho^2$	$2-\rho^2$	$-\rho^2$	1	$7\times 4 \quad =28$
169	$\rho^3+\rho^2-4\rho+1=0$	$-3+2\rho+\rho^2$	$3-\rho-\rho^2$	$-1-\rho$	1	$1\times 4+2\times 12=28$
196	$\rho^3+\rho^2-2\rho-1=0$	ρ	$-\rho-\rho^2$	$-2+\rho^2$	2	$1\times 4+6\times 12=76$

In each case the number of strict proportionality classes (see **4**) of 3×4 matrices is computed. Naturally more than one of these proportionality classes may correspond to the same quaternary form (see **4**) under permutation of the m_i, (or of the P, Q, R, S). On the basis of simple group-theoretical considerations, if more than one proportionality class corresponds to a given quaternary form the correlation is g to 1 where g is the order of a sub-group of the alternating group on P, Q, R, S. This correspondence (of classes of matrices into forms) for the cases in the table is by fours (Klein four-group) or by twelves. For non-abelian cases, if the correspondence is not unique, it is necessarily by fours, as can be easily proved.

14. Units. Consider two 3×4 matrices in a chain or, even more generally, two such matrices having three of their four column vectors respectively equivalent under a unimodular transformation. If such two matrices are *proportional*, they are *strictly proportional*, since the constants of proportionality, being the eigenvalues of a unimodular matrix, have the product 1. Since the eigenvalues are also conjugate algebraic integers, in the conjugate cubic fields, they are either all 1 or are conjugates of a non-trivial unit of the field. From the rational independence of the column vectors, the eigenvalues are all 1 only when the two 3×4 matrices are identical. In fact in the example $(\Delta^2 = 81)$ of fig. 2, there are even closed circuits such as $\phi_0(Q.P.R)\phi_2(R.P.S)\phi_6(S.P.Q)\phi_0$ which produce the trivial unit. (The circuit $\phi_0(Q.P.R)\phi_2(P.S.Q)\phi_8(P.Q.S)\phi_4(Q.R.P)\phi_0$ incidentally produces a non-trivial unit). In each of the five cases treated here, independent (non-trivial) units can be observed. It is clear that as we take an increasing large number of different fields the *a priori* necessity of obtaining different units will increase the complexity of the non-trivial circuits, but no further speculation on this will be attempted at present.

WAYNE UNIVERSITY.

BIBLIOGRAPHY.

[1] P. Bachmann, *Die Arithmetik der quadratischen Formen*, vol. II, Leipzig, 1923, pp. 80-85.

[2] G. Bullig, "Zur Kettenbruchtheorie in Dreidimensionen (Z1)," *Abhandlungen aus dem Mathematischen Seminar der Hanseschen Universitat*, vol. 13, 1940.

[3] H. Hancock, *Development of the Minknowski Geometry of Numbers*, New York, 1923, pp. 378-444.

[4] H. Minkowski, *Diophantische Approximationen*, Leipzig, 1907, pp. 101 ff.

[5] ———, "Zur Theorie der Kettenbruche," *Gesammelte Abhandlungen*, Leipzig, 1911, 1911, pp. 278-292.

ON HYPERBOLIC PARTIAL DIFFERENTIAL EQUATIONS.*

By Philip Hartman and Aurel Wintner.

1. Introduction. In the standard existence and uniqueness theorems for the hyperbolic differential equation

$$(1) \qquad z_{xy} = f(x, y, z, p, q),$$

where $p = z_x$, $q = z_y$, it is assumed that f is continuous in (x, y, z, p, q) and satisfies a uniform Lipschitz condition with respect to (z, p, q). It will be shown (Section **2**) that the assumption of a Lipschitz condition with respect to z can be omitted in these existence theorems, though not in the uniqueness theorems. On the other hand, it will be shown in Section **3** by an example that the Lipschitz condition with respect to (p, q) cannot be omitted and, what is more, that there exist continuous f such that (1) has no solution whatsoever (in a vicinity of a given point of the (x, y)-plane).

The existence theorem to be proved for (1) leads to improvements of some of the results of H. Lewy ([14]; cf. the presentation in [6], pp. 487-508). For example, it will be shown (Section **5**) that if $F(x, y, z, p, q, r, s, t)$ is a function of class C^2, and if the partial differential equation

$$(2) \qquad F(x, y, z, p, q, r, s, t) = 0,$$

where $r = z_{xx}$, $s = z_{xy}$, $t = z_{yy}$, is of hyperbolic type, that is, if

$$(3) \qquad F_s^2 - 4F_rF_t > 0,$$

then initial data (belonging to a z of class C^3) determine a unique C^3-solution of (2). The conclusion, the existence of a solution z of class C^3, is the same as that of Lewy (cf. [14], p. 189, where it is asserted that z, p, q, r, s, t are of class C^1). But Lewy assumes that F is of class C^3 (instead of being, as above, just of class C^2) and he has the somewhat artificial assumption that the initial data belong to a z of class C^4 (instead of belonging, as above, to a z of only class C^3; cf. [14], p. 187, where it is assumed that the initial data of z, p, q, r, s, t are of class C^2). The conclusion of uniqueness is a corresponding improvement of that of Lewy. (Incidentally, his statement of uniqueness involves solutions of a class which is apparently more

* Received December 21, 1951.

restricted than the class C^3, but it turns out that all solutions of class C^3 belong to this restricted class.)

When (2) is of the Monge-Ampère type,

$$A + Br + Cs + Dt + E(rt - s^2) = 0, \tag{4}$$

where A, B, C, D, E are functions of the five variables x, y, z, p, q, and (3) is replaced by

$$C^2 - 4BD + 4AE > 0 \text{ and } E \neq 0, \tag{5}$$

it will be shown (Section **6**) that if A, B, C, D, E are of class C^1, then initial data (belonging to a z of class C^2) determine a unique C^2-solution of (4).

Following Lewy's procedure, the theorems involving (2) will be reduced to one on a system (1) by means of an intermediary system of first order of the type

$$\sum_{k=1}^{n} a_{ik} z^k{}_x = b_i, \qquad 1 \leqq i \leqq m (\leqq n), \tag{6}$$

$$\sum_{k=1}^{n} a_{ik} z^k{}_y = b_i, \qquad m + 1 \leqq i \leqq n, \tag{7}$$

where the given functions a_{ik}, b_i are functions of $x, y, z^1, \cdots, z^n$ satisfying $\det a_{ik} \neq 0$. For the system (6)-(7) it is sufficient to assume that a_{ik}, b_i are of class C^1, in which case initial data belonging to a $z = (z^1, \cdots, z^n)$ of class C^1 determine a unique C^1-solution z which possesses a continuous second mixed derivative $z_{xy} = z_{yx}$; Section **4**. This is an improvement of a theorem of Lewy ([14], pp. 186-187), where it is assumed that a_{ik}, b_i and the initial data are of class C^2, while the assertion is unchanged. The system (6)-(7) can fail to have solutions if it is only assumed that the given functions a_{ik}, b_i are continuous (instead of being of class C^1), with $\det a_{ik} \neq 0$.

It will also be proved that every C^1-solution z of (6)-(7) has a continuous second mixed derivative $z_{xy} = z_{yx}$. This permits a sharpening of the wording of the above-mentioned uniqueness theorem.

In Part II, the results of Part I just outlined will be applied to two problems of differential geometry; first, the problem of embedding into 3-dimensional Euclidean space a positive-definite line element,

$$ds^2 = g_{11} du^2 + 2 g_{12} du dv + g_{22} dv^2, \tag{8}$$

of negative curvature; second, the question of transforming (8) into the Tchebychef form

$$ds^2 = dx^2 + 2 \cos \phi \, dx dy + dy^2, \qquad \phi = \phi(x, y). \tag{9}$$

Part III adapts the method of Part I to the quasi-linear hyperbolic system for the vector $z=(z^1,\cdots,z^n)$,

$$z_x = f z_y + g,$$

where x and y are scalars, $f=f(x,y,z)$ is a matrix and $g=g(x,y,z)$ a vector. The linear case, where $f=f(x,y)$ is independent of z and g is linear in z, was first treated by Holmgren [11] under the assumption that f and g are of class C^2 (and the system is hyperbolic, that is, f has simple elementary divisors and real characteristic numbers). Perron ([16], pp. 562-564) has shown that Holmgren's method remains applicable if g is not linear in z, but is merely of class C^2 in (x,y,z). Friedrichs ([4], pp. 575-578) has proved that the Holmgren-Perron theory is also valid if $f(x,y)$ and $g(x,y,z)$ are only assumed to be of class C^1. The latter result was stated already by Holmgren ([12], pp. 1-2), whose proof however contains a gap.

Friedrichs has extended the results also to the case in which f is allowed to depend on z but both f and g are subject to the classical C^2-restriction ([4], pp. 582-584; cf. also Cinquini-Cibrario [2] and her papers referred to there). It will be shown in Part III below that it is sufficient that $f(x,y,z)$ be of class C^1 only. In particular, all of the results mentioned above appear as consequences of a single theorem. While the method of Holmgren, Perron and Friedrichs is that of the successive approximations, the proof of the general theorem is based in Part III on a combination of the process of successive approximations and the method of equicontinuous functions, as used in Part I.

Part I.

2. The system (1). A function $z=z(x,y)$ will be said to be of class C^* if $z(x,y)$ is of class C^1 and possesses a continuous second mixed derivative $z_{xy}=z_{yx}$.

(I) *Let* $z=(z^1,\cdots,z^n)$, $p=(p^1,\cdots,p^n)$, $q=(q^1,\cdots,q^n)$ *and* $f=(f^1,\cdots,f^n)$ *be vectors with* n *components. Let* $f=f(x,y,z,p,q)$ *be continuous and bounded on the product space of the rectangle*

$$R:\ 0\leqq x\leqq a,\qquad 0\leqq y\leqq b \tag{10}$$

and the entire (z,p,q)*-space, and let* f *satisfy a uniform Lipschitz condition with respect to* (p,q) *on this product space. Let* $\sigma(x)$, $\tau(y)$ *be* n*-dimensional vector functions of class* C^1 *on the respective intervals* $0\leqq x\leqq a$, $0\leqq y\leqq b$,

and satisfy $\sigma(0)=\tau(0)$. *Then there exists on* (10) *a function* $z=z(x,y)$ *of class* C^* *satisfying* (1) *and*

$$z(x,0)=\sigma(x) \text{ and } z(0,y)=\tau(y). \tag{11}$$

It will be clear from the proof that if f is not bounded, then it is possible to find an α, where $0<\alpha\leqq a$, such that the solution $z=z(x,y)$ exists on the square $0\leqq x\leqq\alpha$, $0\leqq y\leqq\alpha$. For example, if the absolute value of the components of $\sigma(x)$, $\sigma'(x)$, $\tau(y)$, $\tau'(y)$ are bounded by a number M, and if $C>0$ is a bound of $|f|$ on the product space of (10) and the $3n$-dimensional parallelopiped $|z^i|\leqq 4M$, $|p^i|\leqq 4M$, $|q^i|\leqq 4M$, where $i=1,2,\cdots,n$, then α can be chosen to be min $(a,b,3M/C,M^{\frac{1}{2}}/C^{\frac{1}{2}})$. In this case, the absolute value of the components of a solution z and its partial derivatives $p=z_x$, $q=z_y$ do not exceed $4M$.

In the simple case when $f(x,y,z,p,q)$ is independent of the derivatives p and q, the assertion (I) reduces to a theorem of Montel [15], p. 282. His proof, as well as that of (I) below, is adapted from the standard proof of the Peano existence theorem in the theory of ordinary differential equations.

Remark. It should be noted that, in the wording of (I), the specification of precisely the class C^* is essential. Of course, every C^1-solution of (1) is a solution of class C^*. On the other hand, the example $z_{xy}=0$ shows that differential equations of type (1) can have solutions which are not of class C^1. In fact, such a solution is $F(x)+G(y)$ if F, G possess first derivatives having discontinuities (which are, of course, of the second kind by necessity).

Actually (1) need not possess any solution of class C^2 if f satisfies the assumptions of (I). In fact, if $f(x,y)$ is a continuous function in a vicinity of $(x,y)=(0,0)$, it is clear that the second derivative ϕ_{xx} (or ϕ_{yy}) of the function

$$\phi(x,y)=\int_0^x\int_0^y f(\alpha,\beta)\,d\alpha d\beta$$

will not in general exist; so that $\phi(x,y)$ fails to be of class C^2 (though it is, of course, of class C^*). In addition, $-\phi(x,y)$ will not become of the form $F(x)+G(y)$ if $f(x,y)\neq 0$. Hence, if $f(x,y)$ is just continuous, then for no choice of $F(x)$ and $G(y)$ need the sum $\phi+F+G$ become of class C^2. Since every solution $z(x,y)$ of (1) is such a sum if $f=f(x,y)$, the assertion follows.

Proof of (I). For a fixed m, let $R_{jk}=R_{jk}{}^m$, where $j,k=0,\cdots,m-1$, denote the rectangle

$$(12) \qquad R_{jk} = R_{jk}{}^m\colon\ ja/m \leqq x < (j+1)a/m,\ kb/m \leqq y < (k+1)b/m.$$

Three vector functions $Z = Z_m(x, y)$, $P = P_m(x, y)$, $Q = Q_m(x, y)$ will be defined on R with the properties that Z, P, Q are constant on R_{jk} and that if

$$(13) \qquad z = z_m(x, y) = \sigma(x) + \tau(y) - \sigma(0) + \int_0^x \int_0^y f(u, v, Z(u, v), P(u, v), Q(u, v))\, du dv,$$

then, on R_{jk},

$$(14) \qquad Z(x, y) = z(ja/m, kb/m),$$

$$(15) \qquad P(x, y) = p(ja/m, kb/m), \text{ where } p(x, y) = \lim_{\alpha \to -0} z_x(x + \alpha, y),$$

$$(16) \qquad Q(x, y) = q(ja/m, kb/m), \text{ where } q(x, y) = \lim_{\beta \to -0} z_y(x, y + \beta)$$

(if either $x = 0$ or $y = 0$, the limits defining p and q in (15) and (16) are meant to be replaced by limits as $\alpha \to +0$ and $\beta \to +0$).

In order to obtain the functions Z, P, Q, first put $Z(x, y) = \sigma(0) = \tau(0)$, $P(x, y) = \sigma_x(0)$, $Q(x, y) = \tau_y(0)$ on R_{00}. Then (13) defines $z(x, y)$ as a function of class C^1 on the closure of R_{00}, and the analogues of (14), (15), (16) hold on $R_{jk} = R_{00}$. Define Z, P, Q on R_{10}, R_{01}, R_{11} by (14), (15), (16), respectively, and define z on the closure of these rectangles by (13). Repeat this procedure, defining Z, P, Q on R_{02}, R_{21}, R_{22}, R_{12}, R_{02} by (14), (15, (16) and z on the closures of these rectangles by (13). Continuing in this fashion, the desired functions Z, P, Q, z are defined on R. (In the last step, where either j or k is $m - 1$, the functions Z, P, Q can be defined on the closure of R_{jk}.)

If none of the functions $|\sigma(x)|$, $|\tau(y)|$, $|\sigma'(x)|$, $|\tau'(y)|$, $|f|$ exceeds the number M, then $|z^i|$, $|p^i|$, $|q^i|$ do not exceed $3M + Mab$, $M + Mb$, $M + Ma$, respectively, for $i = 1, \cdots, n$. In particular, bounds for z, p, q can be chosen independently of m. Hence there exists a subsequence of the integers with the property that the corresponding subsequences of the sequences $z_1, z_2, \cdots, p_1, p_2, \cdots$, and $q_1, q_2, \cdots$ converge on a dense set of points of R. Let $z_1, z_2, \cdots; p_1, p_2, \cdots; q_1, q_2, \cdots$ denote these subsequences. It is clear from (13) and the boundedness of f that $z_1, z_2, \cdots$ is equicontinuous on R. Consequently, $z_1, z_2, \cdots$ converges uniformly to a continuous function $z(x, y)$ on R. The relations (14) show that $Z_1, Z_2, \cdots$ also converges uniformly to $z(x, y)$ on R.

It will be shown that the sequences $p_1, p_2, \cdots$ and $q_1, q_2, \cdots$ also converge uniformly to continuous functions on R. For any (scalar or vector)

function h and for any positive number δ, let $w_h = w_h(\delta)$ denote the least upper bound of $|h(x,y) - h(x^*,y^*)|$ for $|x - x^*| \leqq \delta$, $|y - y^*| \leqq \delta$, where $|h|$ denotes the sum of the absolute values of the components of a vector h. In order to prove the existence of uniform continuous limits for $p_1, p_2, \cdots$ and $q_1, q_2, \cdots$, it is sufficient to show that for every $\epsilon > 0$ there exist a $\delta = \delta_\epsilon > 0$ and an integer $N = N_\epsilon$ such that

$$w_h(\delta) < \epsilon \text{ if } h = p_m, q_m, \text{ and } m \geqq N_\epsilon. \tag{17}$$

To this end, note that if (x, y) is not on any of the lines $x = ja/m$ or $y = jb/m$, where $j = 0, 1, \cdots, m$, then, by (13),

$$p(x,y) = \sigma'(x) + \int_0^y f(x, v, Z(x,v), P(x,v), Q(x,v))dv. \tag{18}$$

Hence if (x^*, y^*) is not on any the lines $x = ja/m$ or $y = jb/m$, and if $|x - x^*| \leqq \delta$, $|y - y^*| \leqq \delta$, then

$$|p(x,y) - p(x^*,y^*)| \leqq w_{\sigma'} + M|y - y^*| + b(w + Kw_P + Kw_Q),$$

where

$$\begin{aligned} w = \text{l. u. b. } |f(x, y, Z(x,y), P(x,y), Q(x,y)) \\ - f(x^*, y^*, Z(x^*,y^*), P(x,y), Q(x,y))| \end{aligned} \tag{19}$$

for $|x - x^*| \leqq \delta$, $|y - y^*| \leqq \delta$, and K is a positive constant implicit in the assumption of a uniform Lipschitz condition on f. Clearly,

$$w_p \leqq w_{\sigma'} + M\delta + b(w + Kw_P + Kw_Q). \tag{20}$$

It is similarly seen that

$$w_q \leqq w_{\tau'} + M\delta + a(w + Kw_P + Kw_Q). \tag{21}$$

The relation (15) shows that $w_P(\delta) \leqq w_P(\delta + 2a/m)$, so that $w_P(\delta) \leqq w_P(2\delta)$ if $2a/m \leqq \delta$. For any function h, $w_h(2\delta) \leqq 2w_h(\delta)$; hence $w_P(\delta) \leqq 2w_p(\delta)$ when $2a/m \leqq \delta$. Similarly, $w_Q(\delta) \leqq 2w_q(\delta)$ if m is sufficiently large. Thus (20) and (21) give, for large m,

$$w_p + w_q \leqq (w_{\sigma'} + w_{\tau'} + 2M\delta + (a + b)w)/(1 - 2(a + b)K), \tag{22}$$

provided that $2(a + b)K < 1$. There is no loss of generality in supposing $2(a + b)K < 1$, for otherwise the rectangle R can be divided into a finite number of sufficiently small rectangles, and theorem (I) applied successively to each of these sub-rectangles (in a suitable order).

In view of the definition (19) of w, the existence, for every $\epsilon > 0$, of a

$\delta = \delta_\epsilon$ and an N_ϵ satisfying (17) follows from (22) and the fact that $Z(x,y) - Z_m(x,y)$ tends, as $m \to \infty$, uniformly on R to a continuous function $z(x,y)$. Consequently, $p_1, p_2, \cdots$ and $q_1, q_2, \cdots$ tend uniformly on R to continuous functions, say $p(x,y)$ and $q(x,y)$. The relations (15) and (16) show that $P_m(x,y) \to p(x,y)$ and $Q_m(x,y) \to q(x,y)$ also hold uniformly on R as $m \to \infty$. It follows from (13) that

$$z(x,y) = \sigma(x) + \tau(y) - \sigma(0) + \int_0^x \int_0^y f(u, v, z, p, q)\,du\,dv \tag{23}$$

on R, and from (18) that $z_x(x,y) = p(x,y) - \sigma'(x) + \int_0^y f(x, v, z, p, q)\,dv$ if (x,y) is not on any line $x = ja/m$ or $y = kb/m$, where $j, k = 0, 1, \cdots, m$ and $n = 1, 2, \cdots$. By the continuity of the functions involved, $z_x(x,y) = p(x,y)$ on R. Similarly, $z_y(x,y) = q(x,y)$. In view of (23), this completes the proof of (I).

(II) *Let $f(x,y,z,p,q)$ satisfy the conditions of* (I). *Let $\sigma(x)$, $\tau(y)$ be functions of class C^1 for $0 \leqq x \leqq a$, $0 \leqq y \leqq b$, respectively. Let $\Gamma: x = x(u)$, $y = y(u)$, where $0 \leqq u \leqq 1$, be an arc of class C^1 such that $(x(0), y(0)) = (0,0)$, $(x(1), y(1)) = (a,b)$ and $x'(u) > 0$, $y'(u) > 0$. Then there exists a function $z = z(x,y)$ of class C^* on* (10) *satisfying* (1) *and*

$$z(x(u), y(u)) = \sigma(x(u)) + \tau(y(u)), \qquad 0 \leqq u \leqq 1. \tag{24}$$

The proof of (II) is similar to that of (I) and will be omitted. If f is not bounded, one obtains a solution $z = z(x,y)$ only in some vicinity of the open arc $\Gamma_0: x = x(u)$, $y = y(u)$, where $0 < u < 1$. For example, if the absolute value of the components of $\sigma(x)$, $\sigma'(x)$, $\tau(y)$, $\tau'(y)$ do not exceed M, and if $|f|$ does not exceed $C > 0$ on the product space of (10) and the $3n$-dimensional parallelopiped $|z^i| \leqq 3M$, $|p^i| \leqq 3M$, $|q^i| \leqq 3M$, where $i = 1, \cdots, n$, then there exists a solution of (1) and (24) on the common part of the two (x,y)-sets

$$\max(0, x(u) - \alpha) \leqq x \leqq \min(a, x(u) + \alpha),\ y = y(u),\ 0 \leqq u \leqq 1, \tag{25}$$

$$x = x(u),\ \max(0, y(u) - \alpha) \leqq y \leqq \min(b, y(u) + \alpha),\ 0 \leqq u \leqq 1, \tag{26}$$

where $\alpha = \min(M^{\frac{1}{2}}/C^{\frac{1}{2}}, 2M/C)$.

3. Counter-examples. In this section, it will be shown that

(i) a solution $z = z(x,y)$ of (1) and (11), whose existence is assured by (I), need not be unique; and that

(ii) (I) becomes false if the Lipschitz condition on f with respect to (p,q) is omitted.

Ad (i). Let $n = 1$ in (I), so that z, p, q, f are scalars. Let $f(x,y,z,p,q) = |z|^\lambda$, where $0 < \lambda < 1$, and let $\sigma(x) = 0$, $\tau(y) = 0$. Thus (1) reduces to $z_{xy} = |z|^\lambda$ and the initial conditions (11) to $z(x,0) = 0$, $z(0,y) = 0$. This initial value problem has on the rectangle $R: 0 \leqq x \leqq 1$, $0 \leqq y \leqq 1$, more than one solution of class C^*, since it is satisfied by both $z(x,y) \equiv 0$ and $z(x,y) = \mu(xy)^{1/(1-\lambda)}$, where $\mu = (1-\lambda)^{2/(1-\lambda)}$.

Ad (ii). By a solution of (1) is meant a function of class C^* satisfying (1). Consider the case of a scalar equation (1) of the form $z_{xy} = f(x, z_y) + y$. It is clear that this equation has a solution of class C^* if and only if the first order *partial* differential equation

$$z_x = f(x,z) + y \tag{27}$$

has a continuous solution $z(x,y)$, which then possesses a continuous partial derivative with respect to x.

Let it be granted for the moment that there exist continuous functions $f(x,z)$ on $-1 \leqq x \leqq 1$, $-\infty < z < \infty$, with the property that the *ordinary* initial value problem

$$dz/dx = f(x,z), \qquad z(0) = z_0, \tag{28}$$

has more than one solution for *every* z_0; and that if $z^*(x)$, $z_*(x)$ denote the maximum and minimum solutions of (28), then

$$z^*(2^{-j}) \geqq z_0 + c2^{-2j} \text{ and } z_*(2^{-j}) \leqq z_0 - c2^{-2j} \tag{29}$$

hold for $j = 0, 1, \cdots$ and some positive c which is independent of z_0 and j. It will be shown that, for such an f, (27) possesses no continuous solution whatever in any square $|x| \leqq \epsilon$, $|y| \leqq \epsilon$. Consequently, there exist continuous functions f such that (1) has no solution (of class C^*) in any vicinity of a given point.

Suppose (27) has a solution $z = z(x,y)$ on some square $|x| \leqq \epsilon$, $|y| \leqq \epsilon$. In particular, $z(0,y)$ is continuous for $|y| \leqq \epsilon$. For a fixed y, the function $z = z(x,y)$ satisfies the ordinary initial value problem

$$dz/dx = f(x,z) + y, \qquad z(0) = z(0,y).$$

If y is positive, then $z(x,y)$ is not less than the maximal solution of (28), where $z_0 = z(0,y)$. Hence $z(2^{-j},y) \geqq z(0,y) + c2^{-2j}$ if $2^{-j} \leqq \epsilon$. Similarly, if y is negative, $z(2^{-j},y) \leqq z(0,y) - c2^{-2j}$. These two inequalities imply that $z(x,y)$ is not continuous at $(x,y) = (2^{-j},0)$, since $c > 0$.

Thus, in order to complete the proof of the assertion (ii), it remains to show that there exist functions $f(x,z)$ having the properties specified in connection with (28). It turns out that an $f(x,z)$ defined, in another connection, by van Kampen ([13], p. 151) has the desired properties. His function $f(x,z)$ is defined as follows:

In an (x,z)-plane, consider the curves

$$(30_1) \qquad z = 4n + \alpha(1 - \cos 2\pi x), \text{ where } 0 \leqq x \leqq 1,$$

$$(30_2) \qquad z = 4m + 2 + \beta(1 + \cos 2\pi x), \text{ where } 0 \leqq x \leqq 1$$

depending on the parameters n, m, α, β, where $n, m = 0, \pm 1, \cdots$ and $-1 < \alpha < 1$, $-1 \leqq \beta \leqq 1$. The curves (30_1) belonging to a fixed n and varying α have only the points $(x,z) = (0,4n)$ and $(1,4n)$ in common, at which points the derivative z_x is 0. These points are also on the curve (30_2), where $m = n - 1$ and $\beta = 1$. The curves (30_2) belonging to a fixed m and varying β have only the point $(x,z) = (\frac{1}{2}, 4m+2)$ in common, at which point the derivative z_x is 0. The points just mentioned are the only points common to two curves in the families (30_1)-(30_2); at such a point $z_x = 0$. Through other points (x,z), where $0 \leqq x \leqq 1$, $-\infty < z < \infty$, there passes one and only one curve in the families (30_1)-(30_2).

Let $g(x,z)$, for $0 \leqq x \leqq 1$, $-\infty < z < \infty$, denote the slope z_x of (30_1)-(30_2) at (x,z). Clearly, $g(x,z)$ is a continuous function and satisfies $g(0,z) = g(\frac{1}{2},z) = g(1,z) = 0$. If z_0 is a number satisfying $4n \leqq z_0 < 4n + 4$, where $n = 0, \pm 1, \cdots$, then the ordinary initial value problem $dz/dx = g(x,z)$, $z(0) = z_0$, has a solution $z = z(x)$ which, for $0 \leqq x \leqq \frac{1}{2}$, is (30_2), where $m = n$ and $4n + 2 + 2\beta = z_0$ (so that $-1 \leqq \beta < 1$), and, for $\frac{1}{2} \leqq x \leqq 1$, is $z(x) = 4n + 2 + (1 + \cos 2\pi x)$. Hence, $z(1) = 4n + 4$.

Define $f(x,z)$ for $-1 \leqq x \leqq 1$, $-\infty < z < \infty$ as follows:

$$f(-x,z) = f(x,z), f(0,z) = 0, \text{ and } f(x,z) = 2^{-j}g(2^j(x - 2^{-j}), 2^{2j}z)$$

if $2^{-j} \leqq x \leqq 2^{-j+1}$, $-\infty < z < \infty$ and $j = 1, 2, \cdots$.

Since $g(0,z) = g(1,z) = 0$, the function $f(x,z)$ is continuous and satisfies $f(2^{-j},z) = 0$ for $j = 0, 1, 2, \cdots$. If $2^{-k} \leqq x \leqq 2^{-k+1}$, the equation $dz/dx = f(x,z)$ is equivalent to $du/dt = g(t,u)$, where $u = 2^{2k}z$ and $t = 2^k(x - 2^{-k})$. Hence it is seen from the above description of certain solu-

tions of the latter differential equation that through the point $(x, z) = (2^{-k}, z_k)$, where $2^{-2k+2}m \leqq z_k < 2^{-2k+2}(m+1)$ and $m = 0, \pm 1, \cdots$, there is a solution $z(x)$ of $dz/dx = f(x, z)$ satisfying $z(2^{-k+1}) = (m+1)2^{-2k+2}$. Hence this solution can be chosen so as to satisfy $z(2^{-k+2}) = (r+1)2^{-2k+4}$ or $z(2^{-k+2}) = (r+2)2^{-2k+4}$ according as m is of the form $4r$, $4r+1$, $4r+2$ or is of the form $4r+3$. Thus, in any case, $z(2^{-k+2}) \geqq z(2^{-k}) + 2^{-2k+2}$ or, if $k-2$ is denoted by k, simply $z(2^{-k}) \geqq z(2^{-k-2}) + 2^{-2-k}$.

If $z^* = z^*(x)$ is the maximal solution of the initial value problem (28), it is clear that $z^*(2^{-k-2}) \geqq z^*(0)$; hence $z^*(2^{-k}) \geqq z^*(0) + 2^{-2k-2}$. This is the first inequality in (19), with $c = 2^{-2} > 0$. It is proved in the same way that the minimal solution $z_*(x)$ of (28) satisfies the second inequality in (29). This completes the proof of (ii).

4. The system (6)-(7). The existence theorem for the system (6)-(7) will be reduced to that of a system (1).

(III) *Let $a_{ik} = a_{ik}(x, y, z)$, $b_i = b_i(x, y, z)$, where $i, k = 1, \cdots, n$, be functions of class C^1 on the product space of the (x, y)-rectangle* (10) *and the n-dimensional z-space $(-\infty < z^1 < \infty, \cdots, -\infty < z^n < \infty)$, and let*

$$\det a_{ik} \neq 0. \tag{31}$$

Let the arc Γ and the functions $\sigma(x)$, $\tau(y)$ satisfy the conditions of (II). *Then there is a number $a > 0$, such that, on the common part of the sets* (25)-(26), *there exists a $z = z(x, y)$ of class C^* satisfying* (6)-(7) *and* (24).

It will be clear from the proof that one can also obtain the existence of solutions of initial value problems of the type (6)-(7) and (11) on a sufficiently small square $0 \leqq x \leqq a$, $0 \leqq y \leqq a$.

It turns out that the situation in (III) is similar to that in (I) and (II) in that every C^1-solution of (6)-(7) is of class C^*.

(iii) *Let a_{ik}, b_i satisfy the conditions of* (III), *and let the vector $z = z(x, y)$ be a function of class C^1 satisfying* (6)-(7) *on some domain. Then $z(x, y)$ is of class C^*.*

In contrast to (I) or (II), there is a uniqueness theorem corresponding to (III) which does not involve any additional hypothesis.

(III bis) *Under the conditions of* (III), *there cannot exist two distinct C^1-solutions $z = z(x, y)$ of* (6)-(7) *and* (24) *on the common part of* (25) *and* (26), *for any $a > 0$.*

Lewy ([14], pp. 186-187) has proved a uniqueness theorem for the case in which the coefficient functions a_{ik} and b_i, and the initial data Γ, $\sigma(x)$ and $\tau(y)$ are of class C^2, and the class of solutions considered is the class C^*. In view of (iii), Lewy's restriction to the class of C^*-solutions is not more severe than the restriction in (III bis) to the class of C^1-solutions.

Both the existence theorem (III) and the uniqueness theorem (III bis) become false if the assumption that a_{ik}, b_i are of class C^1 is lightened to the assumption that they are continuous. As to (III), this is seen from the example (27) above, which is of type (6)-(7), where $m=n=1$. (This remark concerning the impossibility of reducing the assumptions on a_{ik}, b_i in (III) is valid if the system (6)-(7) is of the type considered by Lewy ([14], pp. 186-187), where a_{ik}, b_i do not depend on the independent variables x, y. For, under suitable initial conditions, (27) is equivalent to a system of type (6)-(7), namely to $x_u=1$, $z_u=f(x,z)+y$, $y_v=1$, where x, y, z are the dependent, and u, v the independent, variables and $n=3$, $m=2$.)

As to (III bis), the equation $z_x=|z|^\lambda$, where $0<\lambda<1$, which is of type (6)-(7), with $m=n=1$, has the trivial solution $z(x,y)=0$, as well as a solution given by $z(x,y)=[(1-\lambda)(x-y)]^{1/(1-\lambda)}$ or $z(x,y)=0$ according as $x\geqq y$ or $x<y$. Both of these solutions satisfy (24) if Γ is the line $x=y$, and $\sigma(x)=0$, $\tau(y)=0$.

Proof of (III). If $z=z(x,y)$ is of class C^* and satisfies (6)-(7), let (6), (7) be differentiated with respect to y, x, respectively. In view of (31), the resulting set of n equations can be solved for the n components of $z_{xy}=z_{yx}$. This leads to a system of equations (1) in which each component of f is of the form $U+V+W$, where U is a quadratic form in the components of z_x and z_y with coefficients which are continuous functions of x, y, z, V is a corresponding linear form, and W is a continuous function of x, y, z. Conversely, if $z=z(x,y)$ is any solution of class C^* of this particular system (1), then $z=z(x,y)$ is also a solution of (6)-(7). Consequently, (III) follows from (II) and the remark made immediately after (II).

This proof shows that (6)-(7) is equivalent to a system (1) if only solutions of class C^* are considered. In view of (iii), the same is true if solutions of class C^1 are considered.

Proof of (iii). Let (x_0,y_0) be a point of (10), and let $\alpha\neq 0$, $\beta\neq 0$ be number such that $z(x,y)$ is defined on the rectangle with the vertices $P_1=(x_0,y_0)$, $P_2=(x_0+\alpha,y_0)$, $P_3=(x_0+\alpha,y_0+\beta)$, $P_4=(x_0,y_0+\beta)$. Let (6) be integrated with respect to x from x_0 to $x_0+\alpha$ (along P_1P_2). Then an integration by parts shows that

$$(32)\quad \sum_{k=1}^{n} a_{ik}(P_1)[z^k]_1^2 = -\sum_{k=1}^{n}[a_{ik}]_1^2 z^k(P_2) + \int_{P_1P_2} \sum_{k=1}^{n} a_{ikx} z^k dx + \int_{P_1P_2} b_i dx,$$

where $a_{ikx} = \partial a_{ik}/\partial x + \sum_{j=1}^{n} (\partial a_{ik}/\partial z^j) z_x{}^j$ and, for any function h, the bracket is defined by $[h]_1^2 = h(P_2) - h(P_1)$. Similarly, an integration of (6) from P_4 to P_3 gives

$$(33)\quad \sum_{k=1}^{n} a_{ik}(P_4)[z^k]_4^3 = -\sum_{k=1}^{n}[a_{ik}]_4^3 z^k(P_3) + \int_{P_4P_3} \sum_{k=1}^{n} a_{ikx} z^k dx + \int_{P_4P_3} b_i dx,$$

where the left side can be written as

$$\sum_{k=1}^{n} a_{ik}(P_1)[z^k]_4^3 + \sum_{k=1}^{n}[a_{ik}]_1^4[z^k]_4^3.$$

If $z^k(P_2)$ on the right side of (32) is expanded to $z^k(P_3) - [z^k]_2^3$, then subtraction of (32) from (33) gives

$$(34)\quad \sum_{k=1}^{n} a_{ik}(P_1)[z^k] = -\sum_{k=1}^{n}[a_{ik}]z^k(P_3) + \Big(\int_{P_4P_3} - \int_{P_1P_2}\Big) a_{ikx} z\, dx$$
$$-\sum_{k=1}^{n}[a_{ik}]_1^2[z\ ]_2^3 - \sum_{k=1}^{n}[a_{ik}]_1^4[z^k]_4^3 + \Big(\int_{P_4P_3} - \int_{P_1P_2}\Big) b_i dx,$$

where $[h] = h(P_1) - h(P_2) + h(P_3) - h(P_4)$, for any function h, so that $[h] = [h]_4^3 - [h]_1^2 = [h]_2^3 - [h]_1^4$.

It will be shown that if the right side of (34) is divided by $\alpha\beta$, then the quotient tends to a limit if first $\alpha \to 0$ and then $\beta \to 0$. This is clear for the last three terms of (34). As to the first two terms on the right of (34), the corresponding quotient has, as $\alpha \to 0$, the limit

$$\beta^{-1}\{\sum_{k=1}^{n}[a_{ikx}]_1^4 z^k(P_4) + \sum_{k=1}^{n}[a_{ikx}z^k]_1^4\},\ \text{i. e.,}\ \beta^{-1}\sum_{k=1}^{n} a_{ikx}(P_1)[z^k]_1^4.$$

But this expression tends to a limit as $\beta \to 0$.

For the purpose of combining equations (34), which resulted from (6), with the corresponding equations which result from (7), it will be verified that the ratio of the right side of (34) to $\alpha\beta$ also tends to a limit if $\beta \to 0$ and then $\alpha \to 0$. To this end, add and subtract the term

$$(35)\qquad -\Big(\int_{P_3P_2} - \int_{P_1P_4}\Big)\sum_{k=1}^{n} a_{iky} z^k dy$$

on the right side of (34). This term, when combined with the first term

on the right of (34), behaves as did the first two terms of (34). The ratio of the last three terms of (34) to $\alpha\beta$ clearly has a limit as $\beta \to 0$ and then $\alpha \to 0$.

Thus there remains to consider the difference of (35) and the second term of (34). This difference is the line integral

$$\int -\sum_{k=1}^{n} z^k (a_{ikx} dx + a_{iky} dy)$$

around the oriented rectangle $P_1P_2P_3P_4$. If the derivatives a_{ikx} and a_{iky} are of class C^1, then Green's formula shows that this line integral is

$$\int\int \sum_{k=1}^{n} (a_{ikx} z_y^k - a_{iky} z_x^k)\, dx dy,$$

where the double integral is taken over the rectangle $P_1P_2P_3P_4$. But the Lemma in [8], p. 761, shows that this reduction is also valid in the present case. Since the ratio of this double integral to $\alpha\beta$ tends to a limit as $(\alpha, \beta) \to (0, 0)$, this completes the proof of the statement made before (35).

Let (34′) denote the set of equations which results from (7) in the same way as (34) did from (6). Then, by (31), the system of linear equations (34), (34′) can be solved for the components of $[z]$, and so the preceding considerations show that

$$\lim_{\alpha \to 0} \lim_{\beta \to 0} (\alpha\beta)^{-1}[z] = \lim_{\alpha \to 0} \alpha^{-1} (z_y(P_2) - z_y(P_1)) = z_{xy}(P_1)$$

exists and is continuous. This proves (iii).

Proof of (III bis). Suppose that (6)-(7) and (24) have two C^1-solutions, say $z = z(x, y)$ and $z = z(x, y) + \zeta(x, y)$, on the common part of (25) and (26), for some $a > 0$. The difference, ζ, of these solutions vanishes on Γ.

If, as above, $a_{ik} = a_{ik}(x, y, z)$, $b_i = b_i(x, y, z)$, then it follows from (6) that, if $1 \leqq i \leqq m$,

$$\sum_{k=1}^{n} a_{ik} \zeta_x^k + \sum_{k=1}^{n} (z_x^k + \zeta_x^k)(a_{ik}(x, y, z + \zeta) - a_{ik}) = b_i(x, y, z + \zeta) - b_i.$$

Since a_{ik}, b_i are of class C^1, this implies that

$$\sum_{k=1}^{n} a_{ik} \zeta_x^k = O(\sum_{k=1}^{n} | \zeta^k |),$$

where the constant implicit in the O-term depends only on the partial derivatives of $z + \zeta$, a_{ik} and b_i. If (x, y) is in the common part of the sets (25)-(26), let the last relation be integrated with respect to x from the point

$(x(u), y(u))$ on Γ, where $y = y(u)$, to the point (x, y). An integration by parts shows that

$$\sum_{k=1}^{n} a_{ik}\zeta^k = \int_{x(u)}^{x} \sum_{k=1}^{n} a_{ikx}\zeta^k dx + \int_{x(u)}^{x} O(\sum_{k=1}^{n} | \zeta^k |) dx,$$

since ζ vanishes on Γ. Consequently,

$$\sum_{k=1}^{n} a_{ik}\zeta^k = O(\int_{x(u)}^{x} \sum_{k=1}^{n} | \zeta^k | dx) \text{ for } 1 \leqq i \leqq m.$$

Similarly, (7) implies that

$$\sum_{k=1}^{n} a_{ik}\zeta^k = O(\int_{y(v)}^{y} \sum_{k=1}^{n} | \zeta^k | dx) \text{ for } m+1 \leqq i \leqq n,$$

where $(x(v), y(v))$ is the point of Γ for which $x(v) = x$. Since (31) shows that the system of n linear equations represented by the last two formula lines can be solved for the n components of ζ, it is seen that

$$| \zeta(x, y) | = nO(\int_{x(u)}^{x} | \zeta | dx + \int_{y(v)}^{y} | \zeta | dy), \text{ where } | \zeta | = \sum_{k=1}^{n} | \zeta^k |.$$

Hence if μ is the maximum of $| \zeta |$ on the common part of (25) and (26), then $\mu = 2naO(\mu)$.

Since the constant implicit in the last O depends only on the functions z, ζ, a_{ik}, b_i and their partial derivatives, it follows that $\mu = 0$ if a is sufficiently small. In view of the definition of μ, this means that $\zeta(x, y) \equiv 0$, so that the solution of (6)-(7) and (24) is unique on the common part of (25) and (26), if a is sufficiently small. Actually, the last proviso concerning the smallness of a is superfluous, for it is only necessary to divide the common part of (25) and (26) into a finite number of suitable parts and to apply the arguments just used successively on these parts, arranged in a convenient order.

5. The hyperbolic equation (2). The following theorem is the improved version, mentioned in Section **1**, of the corresponding result of Lewy ([14], pp. 187-191).

(IV) *Let* $\Gamma : x = x(u)$, $y = y(u)$, *where* $0 \leqq u \leqq 1$, *be an arc of class* C^3, *and* Γ_0 *the open part of* Γ, $0 < u < 1$. *For* $0 \leqq u \leqq 1$, *let* $z = z(u)$

be of class C^3, and $p = p(u)$, $q = q(u)$ of class C^2, finally $r = r(u)$, $s = s(u)$, $t = t(u)$ of class C^1. In an 8-dimensional open set containing the points $(x(u), y(u), z(u), p(u), q(u), r(u), s(u), t(u))$, where $0 \leqq u \leqq 1$, let $F(x, y, z, p, q, r, s, t)$ be a function of class C^2 satisfying (3). For $0 \leqq u \leqq 1$, let the functions $x = x(u), \cdots, t = t(u)$ satisfy the relations (2),

(36) $z' = px' + qy'$, (37) $p' = rx' + sy'$, (38) $q' = sx' + ty'$,

and

$$F_r y'^2 - F_s x'y' + F_t x'^2 \neq 0, \tag{39}$$

where $' = d/du$. Then, in a vicinity of Γ_0, there exists one and only one function $z = z(x, y)$ of class C^3 which satisfies (2) and the initial conditions $z(x(u), y(u)) = z(u), \cdots, z_{yy}(x(u), y(u)) = t(u)$, $0 \leqq u \leqq 1$.

Although the statement of (IV) is rather awkward, (IV) is simply the analogue of (II) when z is a scalar. The arc Γ plays the same role in both theorems. The conditions $x'(u) > 0$, $y'(u) > 0$ in (II), which mean that Γ is not in a characteristic direction at any point, are replaced by (39). The initial conditions in (II) are given by (24) in terms of two arbitrary functions $\sigma(x)$, $\tau(y)$ or, equivalently, by the derivatives $\sigma'(x)$, $\tau'(y)$ and quadratures. The same is true in (IV), where, if $p(u)$ and $q(u)$ are given, $z = z(u)$ is determined by a quadrature of (36), and $r = r(u)$, $s = s(u)$, $t = t(u)$ are determined by (2), (37), and (38). In fact, if x, y, z, p, q are considered as functions of u, and r, s, t as unknowns in the equations (2), (37), (38), then (39) expresses the non-vanishing of a Jacobian, thus assuring that r, s, t are uniquely determined as functions of u (if (2), (37), (38) are satisfied for at least one value of u).

The proof of (IV) will not be given, since it becomes identical with that of Lewy's, if use is made of (III), (III bis) above instead of his theorem on systems of the type (6)-(7). (Another reason for not giving the proof of (IV) is that a similar proof will be given in detail for the case of a Monge-Ampère equation in the next section.)

6. The Monge-Ampère equation (4). As is well known (cf., e. g., [5], pp. 42-43), if the hyperbolic differential equation (2) is of the Monge-Ampère type, then the differential equations for (x, y, z, p, q, r, s, t), which define a characteristic strip of (2), can be replaced by differential equations which involve only (x, y, z, p, q). This formal simplification permits a reduction of the degree of differentiability in the assumptions and assertions of (IV) when (2) is of the form (4). This is the content of the following theorem.

(V) *Let $\Gamma: x = x(u)$, $y = y(u)$, where $0 \leqq u \leqq 1$, be an arc of class C^2, and Γ_0 the open part of Γ, $0 < u < 1$. For $0 \leqq u \leqq 1$, let $z = z(u)$ be a function of class C^2, and let $p = p(u)$, $q = q(u)$ be functions of class C^1. In a 5-dimensional open set containing $(x(u), \cdots, q(u))$, where $0 \leqq u \leqq 1$, let A, B, C, D, E be functions of x, y, z, p, q, of class C^1 satisfying (5), and let $\lambda_1 = \lambda_1(x, y, z, p, q)$, $\lambda_2 = \lambda_2(x, y, z, p, q)$ denote the (real, distinct) roots of the quadratic equation*

$$\lambda^2 + C\lambda + BD - AE = 0, \tag{40}$$

so that λ_1, λ_2 can be chosen to be of class C^1.

For $0 \leqq u \leqq 1$, let $x = x(u), \cdots, q = q(u)$ satisfy (36) and

$$(\lambda_2 x' + By' + Eq')(Dx' + \lambda_1 y' + Ep') \neq 0. \tag{41}$$

Then, in a vicinity of Γ_0, there exists one and only one function $z = z(x, y)$ of class C^2 satisfying (4) and the initial conditions $z(x(u), y(u)) = z(u)$, $z_x(x(u), y(u)) = p(u)$ and $z_y(x(u), y(u)) = q(u)$ for $0 \leqq u \leqq 1$.

Proof of (V). Consider the system of partial differential equations

$$(42_1) \quad \lambda_1 x_u + By_u + Eq_u = 0, \qquad (42_2) \quad Dx_u + \lambda_2 y_u + Ep_u = 0,$$

$$(42_3) \quad \lambda_2 x_v + By_v + Eq_v = 0, \qquad (42_4) \quad Dx_v + \lambda_1 y_v + Ep_v = 0,$$

$$(42_5) \qquad px_v + qy_v - z_v = 0$$

in the dependent variables $(z^1, z^2, \cdots, z^5) = (x, y, z, p, q)$ and the independent variables u, v (which do not occur explicitly). This system is of the type (6)-(7), where $n = 5$ and $m = 2$. It is readily verified that the determinant of the matrix of the coefficient functions is

$$-E^2(\lambda_1 - \lambda_2)^2 = -E^2(C^2 - 4BD + 4AE)^2,$$

which, by (5), is not 0. The coefficient functions of (42) are of class C^1. Consequently, (III) is applicable if suitable initial conditions are given.

In order to choose such conditions, let Γ^* be the segment $v = u$, $0 \leqq u \leqq 1$, in the (u, v)-plane. Then, corresponding to the initial functions $x = x(u), \cdots, q = q(u)$, there is an a, $0 < a < 1$, such that in the closed hexagonal region H bounded by the six line segments

$$(v = 0, 0 \leqq u \leqq a), \quad (v = u - a, a \leqq u \leqq 1), \quad (u = a, 1 - a \leqq v \leqq 1),$$

$$(v = 1, 0 - a \leqq v \leqq 1), \ (v = u + a, 0 \leqq u \leqq 1 - a), \ (u = 0, 0 \leqq u \leqq a),$$

there exist five functions $x = x(u, v), \cdots, q = q(u, v)$ of class C^* which

satisfy (42_1)-(42_5), and which reduce to the given functions $x(u), \cdots, q(u)$ when $v = u$, $0 \leqq u \leqq 1$.

It will be shown that, in the region H, the functions x, y, p, q of (u, v) satisfy the identity

$$(43) \qquad p_u x_v + q_u y_v = p_v x_u + q_v y_u.$$

To this end, multiply (42_1), (42_4) by x_v, x_u, respectively, and subtract the resulting two equations. This gives

$$\lambda_2 y_u x_v - \lambda_1 y_v x_u + E(p_u x_v - p_v x_u) = 0.$$

Similarly, (42_1) and (42_3) lead to

$$\lambda_2 y_u x_v - \lambda_1 y_v x_u + E(q_v y_u - q_u y_v) = 0.$$

Since $E \neq 0$ by (5), the last two relations show that

$$(44) \qquad p_u x_v - p_v x_u = q_v y_u - q_u y_v,$$

and this is equivalent to (43).

The following analogue of (42_5) will now be verified:

$$(45) \qquad p x_u + q y_u - z_u = 0.$$

If $v = u$, then (45) follows from (42_5) and (36). Hence, (45) will be proved on H it is shown that the right side of (45) has, with respect to v, a partial derivative which vanishes identically. Since $x, \cdots, q$ are of class C^*, the right side of (45) has, with respect to v, a partial derivative which is equal to

$$p_v x_u + q_v y_u + p x_{uv} + q y_{uv} - z_{uv}.$$

A differentiation of (42_5) with respect to u and the identity (43) show that the last expression is identically 0 (since $x_{uv} = x_{vu}$, $y_{uv} = y_{vu}$, $z_{uv} = z_{vu}$). This proves (45).

The Jacobian of x, y with respect to u, v does not vanish on Γ^*, that is,

$$(46) \qquad x_v y_u - x_u y_v \neq 0$$

for $v = u$ and $0 \leqq u < 1$. First, if (46) does not hold at some point of Γ^*, then, by (42_1)-(42_4) and $E \neq 0$, the ratios $x_u : x_v$, $y_u : y_v$, $p_u : p_v$, $q_u : q_v$ are identical. But then $x_u : x'$, $y_u : y'$, $p_u : p'$, $q_u : q'$ are identical, where $'$ denotes differentiation in any direction, say along $v = u$. This contradicts (41). Hence (46) holds in the region H if a is sufficiently small.

In view of (46), the linear equations

$$(47) \qquad p_u = r x_u + s y_u, \qquad p_v = r x_v + s y_v$$

determine a pair of continuous functions $r = r(u,v)$, $s = s(u,v)$ on H. Similarly,

(48) $$q_u = sx_u + ty_u, \qquad q_v = sx_v + ty_v$$

determine a pair of continuous functions $s = s(u,v)$, $t = (u,v)$ on H. It follows from (44) and (46) that the two definitions (47), (48) of $s = s(u,v)$ are identical.

That the functions $r(u,v), \cdots, t(u,v)$ satisfy (4) follows by standard procedures, which will be indicated here for the sake of completeness. Multiply (42_1), (42_2) by y_u, x_u, respectively, and add. Since (40) shows that $\lambda_1 + \lambda_2 = -C$, the result is

$$By_u^2 - Cx_uy_u + Dx_u^2 + E(p_ux_u + q_uy_u) = 0,$$

or, in view of (47),

(49) $$By_u^2 - Cx_uy_u + Dx_u^2 + E(rx_u^2 + 2sx_uy_u + ty_u^2) = 0.$$

Multiply (42_1) by $Dy_u - Eq_u - \lambda_2 x_u$, and (42_2) by $Bx_u - E_u - \lambda_1 y_u$, and add. Using the relations $\lambda_1 + \lambda_2 = -C$ and $\lambda_1\lambda_2 = BD - AE$, one obtains

$$A(x_u^2 + y_u^2) + (B - D)(p_ux_u - q_uy_u) + C(p_uy_u + q_ux_u) - E(p_u^2 + q_u^2) = 0,$$

after dividing by $E \neq 0$. If p_u, q_u are eliminated from this relation by the use of (47), (48), it follows that

$$(A + Cs)(x_u^2 + y_u^2) + (B - D)(rx_u^2 - ty_u^2) + Cx_uy_u(r + t)$$
$$- E[s^2(x_u^2 + y_u^2) + r^2x_u^2 + t^2y_u^2 + 2(r + t)sx_uy_u] = 0.$$

If (49) is multiplied by $r + t$ and the product added to the last equation, it is seen that

$$\{A + Br + Cs + Dt + E(rt - s^2)\}(x_u^2 + y_u^2) = 0.$$

Since, by (46), the last factor does not vanish, (4) is satisfied on H.

It follows from (46) that in a 2-dimensional vicinity of the open segment $\Gamma_0^*: v = u$, $0 < u < 1$, the transformation $x = x(u,v)$, $y = y(u,v)$ from (u,v) to (x,y) has an inverse (of class C^1). Hence the existence statement in (V) follows.

The uniqueness statement follows from (III bis) since, as is well known, every C^2-solution of (4) is built up of characteristics; cf. [5], pp. 43-44.

Part II.

7. **Embedding a hyperbolic ds^2.** Let $(g_{ik}) = (g_{ik}(u,v))$ be a binary, symmetric, positive-definite matrix of class C^2 in a vicinity of the point $(u,v) = (0,0)$. The line element (8) is called hyperbolic if the curvature $K = K(u,v)$ belonging to (8) is negative. The problem of the existence of a surface $X = X(u,v)$, in 3-dimensional Euclidean space, $X = (x,y,z)$, such that

$$(50) \qquad |\,dX\,|^2 = g_{11}du^2 + 2g_{12}dudv + g_{22}dv^2,$$

can be reduced to the determination of a function $z = z(u,v)$ such that the line element

$$(51) \qquad ds^2 - dz^2 = (g_{11} - z_u{}^2)du^2 + 2(g_{12} - z_u z_v)dudv + (g_{22} - z_v{}^2)dv^2$$

is positive-definite with a curvature 0, and to the determination of solutions of the equation

$$(52) \qquad dx^2 + dy^2 = ds^2 - dz^2$$

for the functions $x = x(u,v)$, $y = y(u,v)$ (Weingarten; cf. [3], pp. 253-254).

When the coefficients g_{ik} are of class C^2, the problem of determination of a $z(u,v)$ such that (51) has zero curvature leads to a Monge-Ampère equation (4), where $p = z_u$, $q = z_v, \cdots, t = z_{vv}$; cf. e. g. [3], pp. 253-254. The coefficients B, C, D are linear forms in p, q, with coefficients which depend on g_{ik} and their first order partial derivatives. The coefficient E is $-(\det g_{ik})^{\frac{1}{2}} < 0$, and the function A is $Q + K(\det g_{ik})^2$, where K is the curvature of (8) and Q is a quadratic form in p, q with coefficients depending on g_{ik} and their first order partial derivatives; cf. [9], p. 883. Hence, if $p = q = 0$, the right side of the first inequality in (5) is $-4K(\det g_{ik})^{5/2}$, so that (5) holds if (8) is hyperbolic.

Consequently, the existence theorem (V) for a $z = z(u,v)$, with $p(0,0) = q(0,0) = 0$, is applicable when the functions g_{ik} are of class C^3 (so that A, B, C, D, E are of class C^1). The theorem (V) shows that in this case there exist functions $z = z(u,v)$ of class C^2 in a vicinity of $(u,v) = (0,0)$ such that (51) is positive-definite and of curvature 0. A result in [9], p. 882, then implies the existence of functions $x = x(u,v)$, $y = y(u,v)$ of class C^2 in a vicinity of $(u,v) = (0,0)$ satisfying (52). Hence one obtains the following theorem:

(*) *Let $(g_{ik}(u,v))$ be a binary, symmetric positive-definite matrix of class C^3 in a vicinity of $(u,v) = (0,0)$. Then there exists a vector function*

$X = (x(u,v), y(u,v), z(u,v))$ of class C^2 in a vicinity of $(u,v) = (0,0)$ with the property that (50) holds.

Remark. The statement (*) remains true if the classes C^3 and C^2 in the assumption and in the assertion are replaced by C^{n+1} and C^n, respectively, where $n > 2$ is arbitrary.

The truth of this remark follows from the corresponding analogues of (II), (III) and (V), analogues which (n in C^n being at least 3) are supplied by the classical theory of characteristics (and which are clear, of course, from the proofs of the above theorems also).

This theorem contrasts with that of [9], pp. 876-877, where (8) is of the elliptic type ($K > 0$), in which case it is sufficient to assume that the g_{ik} have second order partial derivatives satisfying a Hölder condition (instead of assuming that the g_{ik} are of class C^3) in order to assure the existence of an $X(u,v)$ of class C^2 satisfying (50). In the treatment of the hyperbolic case ($K < 0$), we have blundered in [7], [9]; see [10]. The above result and the remark following (*) fill in the gap created by that blunder.

8. Tchebychef nets. The following theorem on the possibility of transforming (8) into the Tchebychef form (9) will be proved:

(†) *In a vicinity of $(u,v) = (0,0)$, let the coefficients* of

$$ds^2 = g_{11}(u,v)\,du^2 + 2g_{12}(u,v)\,du\,dv + g_{22}(u,v)\,dv^2 \tag{53}$$

be functions of class C^1 and have a non-vanishing determinant. Then there exists, in a vicinity of $(x,y) = (0,0)$, a pair of functions

$$u = u(x,y), \qquad v = v(x,y) \tag{54}$$

which are of class C^1, of non-vanishing Jacobian, and such that (53) *becomes identical with*

$$ds^2 = dx^2 + 2\gamma(x,y)\,dx\,dy + dy^2 \tag{55}$$

by virtue of (54).

As observed by Bianchi ([1], vol. I, p. 157), the possibility of transforming (53), when

$$\det g_{ik} \neq 0, \tag{56}$$

into the normal form (55) follows from an existence theorem of classical type if the functions g_{ik} have continuous partial derivatives of first order satisfying a uniform Lipschitz condition. Thus, the point in (†) is that the

imposition of a Lipschitz restriction is claimed to be superfluous. The role of the assumption (56) is that of assuring the existence of the contravariant representation (g^{ik}) of the tensor (g_{ik}), which, combined with the C^1-character of g_{ik}, means that the Christoffel symbols $\Gamma_{jk}{}^i = \Gamma_{jk}{}^i(u, v)$ exist and are continuous. Hence, it is possible to consider the pair of Servant's partial differential equations

$$(57) \qquad u^i{}_{xy} = -\Gamma_{\alpha\beta}{}^i(u^1, u^2)\, u^\alpha{}_x u^\beta{}_y, \text{ where } (u^1, u^2) = (u, v),$$

which represent the formal condition for a pair of functions (54) transforming (53) into (55); cf. [1], vol. I, pp. 156-157.

If (56) is omitted, the assertion of (†) becomes false even if the coefficient functions of (53) are analytic. For instance, if (53) is chosen to be the square of $g(v)du$, where $g(v)$ is a non-constant, positive, analytic function of v, and if (53) is identical with (55) by virtue of (54), then $gu_x = 1$ and $gu_y = 1$. This implies on the one hand that $u_x = u_y$, which means that u is a function of $x + y$ alone, and on the other hand that $g^{-1}(v)(dx + dy)$ is a complete differential, and so $g^{-1}(v)$, therefore v itself, is a function of $x + y$ alone. Hence the Jacobian of (54) must vanish identically, and so (54) cannot be a transformation.

Proof of (†). Since the $\Gamma_{jk}{}^i$ are continuous, (I) is applicable to (57), which is of type (1) with $(z^1, z^2) = (u^1, u^2)$. Let $\epsilon > 0$, and let $\sigma(x) = (\sigma^1(x), \sigma^2(x))$ and $\tau(y) = (\tau^1(y), \tau^2(y))$ be vectors of class C^1 for $|x| \leqq \epsilon$, $|y| \leqq \epsilon$, respectively, satisfying $\sigma(0) = \tau(0) = 0$. Then (I) implies that, if ϵ is sufficiently small, (57) has on the square $|x| \leqq \epsilon$, $|y| \leqq \epsilon$, a solution $(u(x, y), v(x, y)) = (u^1(x, y), u^2(x, y))$ of class C^* satisfying

$$(58) \qquad u^i(x, 0) = \sigma^i(x), \qquad u^i(0, y) = \tau^i(y), \qquad (i = 1, 2).$$

Let the initial data $\sigma(x)$, $\tau(y)$ be so chosen that

$$(59) \qquad g_{\alpha\beta}(\sigma^1(x), \sigma^2(x))\sigma^{\alpha\prime}(x)\sigma^{\beta\prime}(x) \equiv 1, \qquad |x| \leqq \epsilon, \qquad ({}' = d/dx),$$

$$(60) \qquad g_{\alpha\beta}(\tau^1(y), \tau^2(y))\tau^{\alpha\prime}(y)\tau^{\beta\prime}(y) \equiv 1, \qquad |y| \leqq \epsilon, \qquad ({}' = d/dy),$$

and that

$$(61) \qquad \sigma^{1\prime}(0)\tau^{2\prime}(0) - \sigma^{2\prime}(0)\tau^{1\prime}(0) \neq 0.$$

In view of (58) and (61), the Jacobian of (54) does not vanish at $(x, y) = (0, 0)$ and, therefore, for $|x| \leqq \epsilon$, $|y| \leqq \epsilon$ if ϵ is sufficiently small. It remains to show that (53) is transformed into (55), that is, that

$$(62) \qquad g_{\alpha\beta}(u^1, u^2)u^{\alpha\prime}u^{\beta\prime} \equiv 1, \text{ where } u^i = u^i(x, y)$$

and where the primes denote either $\partial/\partial x$ or $\partial/\partial y$.

This will be deduced from the following geometrical interpretation of (57): If $y = y_0$ is fixed, the vectors $(u^1{}_y(x, y_0), u^2{}_y(x, y_0))$ form a parallel sheaf in the sense of Levi-Civita along the arc $u^1 = u^1(x, y_0)$, $u^2 = u^2(x, y_0)$; cf. [1], vol. II, pp. 806-811. Hence, the squared length $g_{\alpha\beta}(u^1, u^2)u^\alpha{}_y u^\beta{}_y$, where $u^i = u^i(x, y_0)$, is independent of x. But if $x = 0$, it is seen from (58) and (60) that this length is 1. Consequently, (62) holds if the primes denote $\partial/\partial y$. Since it is clear that the same is true if the primes denote $\partial/\partial x$, the proof of (†) is complete.

Part III.

9. Quasi-linear systems. The general theorem announced at the end of Section **1** is as follows:

(VI) *Let f and g in the system of partial differential equations*

$$z_x = f z_y + g, \tag{1}$$

where $z = (z^1, \cdots, z^n)$ is a vector, satisfy the following conditions: $f = f(x, y, z)$ is a matrix function of class C^1 on the (x, y, z)-region

$$(2_1) \quad 0 \leqq x \leqq a, \ |y| \leqq b \qquad\qquad (2_2) \quad |z^1| \leqq c, \cdots, |z^n| \leqq c,$$

and has simple elementary divisors and real characteristic numbers $\lambda_1(x, y, z)$, $\cdots, \lambda_n(x, y, z)$, while $g = g(x, y, z)$ is a vector function which is continuous and possesses continuous partial derivatives with respect to $y, z^1, \cdots, z^n$ on (2_1)-(2_2). Then there exists a positive $a (\leqq a)$ such that, if the constant K is so chosen that

$$|\lambda_1| \leqq K, \cdots, |\lambda_n| \leqq K \text{ on } (2_1)\text{-}(2_2), \tag{3}$$

the system (1) *possesses on the region*

$$0 \leqq x \leqq a, \qquad |y| + Kx \leqq b \tag{4}$$

one and only one C^1-solution $z = z(x, y)$ satisfying

$$z(0, y) = 0. \tag{5}$$

It will be clear from the proof that it is possible to give a (lower) estimate for a in terms of bounds of the elements of the matrix f and their partial derivatives and of the components of g. Theorem (VI) implies that (1) and arbitrary initial C^1-data

$$z(0, y) = z(y), \text{ where } |y| \leqq b \text{ and } |z^i(y)| < c, \quad i = 1, \cdots, n, \tag{5'}$$

have one and only one C^1-solution on a region of the type (4).

7

10. Preliminaries. For a vector (or matrix) v, the symbol $|v|$ will denote the absolute value of the largest component (or element) of v.

The assumptions of (VI) for the matrix $f = f(x, y, z)$ imply that there exists on (2_1)-(2_2) a pair of real matrices $H = H(x, y, z)$, $F = F(x, y, z)$ of class C^1 such that H is non-singular, F is diagonal, and

$$Hf = FH, \tag{6}$$

where $|F| \leqq K$ by (3).

The C^1-character of F (hence of H) can be deduced as follows: The diagonal elements of F are the zeros of the invariant factors of $f - \lambda I$, where I is the unit matrix. When f is of class C^1, the invariant factors of $f - \lambda I$ are polynomials in λ with coefficients which are of class C^1 in $(x, y; z)$. The assumption that the elementary divisors of f are simple means that the invariant factors have no multiple zeros, and so their zeros are of class C^1.

The symbol M will be used to denote a constant (not always the same constant) which is a bound for $|v|$ on (2_1)-(2_2), if v is a combination of f, g, F, H, and their partial derivatives.

In the remainder of this section, all functions will be scalar functions.

Let $F(x, y)$ be a function of class C^1 on

$$0 \leqq x \leqq a, \qquad |y| + Kx \leqq b, \tag{7}$$

and suppose that, on (7),

$$|F| \leqq K. \tag{8}$$

Let $\phi(\xi, \eta)$ be the solution of the ordinary initial value problem

$$d\phi/d\xi + F(\xi, \phi) = 0, \qquad \phi(0, \eta) = \eta, \tag{9}$$

and $\psi = \psi(x, y)$ the solution of the implicit equation

$$\phi(x, \psi) = y. \tag{10}$$

The transformation $(x, y) \rightarrow (\xi, \eta)$,

$$\xi = x, \qquad \eta = \psi(x, y), \tag{11}$$

is defined and of class C^1 on (7), and has the inverse

$$x = \xi, \qquad y = \phi(\xi, \eta) \tag{12}$$

and the non-vanishing Jacobian

$$\partial(\xi, \eta)/\partial(x, y) = \partial\psi/\partial y = \exp \int_0^x F_y(t, \phi(t, \psi(x, y)))\,dt. \tag{13}$$

If $G(x,y)$ is a continuous function on (7), then the partial intial value problem

$$z_x = Fz_y + G, \qquad z(0,y) = 0 \tag{14}$$

is transformed by (11) into

$$z_\xi = G, \qquad z(0,\eta) = 0. \tag{15}$$

The latter problem has the unique solution

$$z = \int_0^\xi G(t, \phi(t,\eta))\,dt, \tag{16}$$

or, in terms of (x,y),

$$z = \int_0^x G(t, \phi(t, \psi(x,y)))\,dt. \tag{17}$$

If G is of class C^1, then (17), defined on (7), is the unique solution of (14). Under the assumption of mere continuity of G, it cannot be verified and, in general, is not true, that (17) is a solution of (14). If, however, G is of the form

$$G = (H_x - FH_y)U + V, \tag{18}$$

where $H = H(x,y)$, $U = U(x,y)$, $V = V(x,y)$ are of class C^1 on (7), then (17) is of class C^1 on (7) and is the unique solution of (14). (This remark is a simple case of Theorem 4.4 in [4], pp. 570-571). In order to verify this remark, note that (15), (16) can be written as

$$z_\xi = H_\xi U + V, \qquad z(0,\eta) = 0, \tag{19}$$

$$z = \int_0^\xi (H_\xi U + V)\,dt, \tag{20}$$

respectively. Let Δh denote the difference between the value of the function h at the points $(\xi, \eta + \delta)$, (ξ, η). Then (20) shows that

$$\Delta z = \int_0^\xi \{(\Delta H)_\xi U + H_\xi^* \Delta U + \Delta V\}\,dt,$$

where the argument of H_ξ^* corresponds to $(\xi, \eta + \delta)$. The first term can be integrated by parts and is

$$(\Delta H)U\,\Big|_0^\xi - \int_0^\xi (\Delta H)U_\xi\,dt.$$

This shows that z has a continuous partial derivative with respect to η,

$$z_\eta = H_\eta U \,|_0^\xi + \int_0^\xi \{-H_\eta U_\xi + H_\xi U_\eta + V_\eta\} dt. \tag{21}$$

It is seen from (20) and (21) that z is of class C^1 with respect to (ξ, η), hence with respect to (x, y). It is also clear that z is the unique solution of (14), (18).

The derivative z_y can be calculated from (13) and (21),

$$z_y = H_y U + \int_0^x \{H_x U_y - H_y U_x + V_y\} \exp\left(\int_t^x F_y d\tau\right) dt, \tag{22}$$

where the argument of H_x, H_y, U_x, U_y, V_y in the integral is $(t, \phi(t, \psi(x, y)))$, that of F_y is $(\tau, \phi(\tau, \psi(x, y)))$ and, in (22), it is assumed that $U(0, y) = 0$.

11. The semi-linear case. In order to obtain certain inequalities for the proof of (VI), the existence statement in the semi-linear case of (VI), where the matrix f does not depend on the vector z, will be considered first. In this case, the matrices H, F in (6) depend only on (x, y).

Remark. In the application of this section to the proof of (VI), the matrix $f(x, y, z)$ will be made a function of (x, y) alone, by letting z be a function, says $z = \zeta(x, y)$, of class C^1. Then H and F also become functions of (x, y) alone. In accordance with the paragraph following (6), the symbol M will again be used to denote an upper bound of $|H|$ or $|F|$. On the other hand, C (not M) will be used as an upper bound of $|H_x|$, $|H_y|$, $|F_x|$ and $|F_y|$, since these partial derivatives depend not only on the functions f, g but also on the partial derivatives of the function $\zeta(x, y)$. In other words, C can be taken to be $M(1 + B)$, where B is an upper bound of $|\zeta_x|$ and $|\zeta_y|$.

Existence proof for the semi-linear case. This proof is a modification of that of Holmgren [11], pp. 3-8 and Perron [16], pp. 562-564. Since it is assumed that f, H, and F depend only on (x, y), the introduction of the new dependent variable

$$u = Hz \tag{23}$$

changes (1), (5) into

$$u_x = F u_y + (H_x - F H_y) H^{-1} u + Hg, \qquad u(0, y) = 0, \tag{24}$$

if use is made of the identities (6) and of the relations $HH^{-1}{}_y = -H_y H^{-1}$, $HH^{-1}{}_x = -H_x H^{-1}$.

Consider the successive approximations $u_0(x,y), u_1(x,y), \cdots$ defined as follows: $u_0(x,y)=0$ and, for $k=1,2,\cdots$,

$$u_{kx}=Fu_{ky}+(H_x-FH_y)H^{-1}u_{k-1}+Hg(x,y,H^{-1}u_{k-1}), \tag{25}$$

$$u_k(0,y)=0. \tag{25 bis}$$

Since F is a diagonal matrix, it is easily verified that each of the n initial value problems for the components of u_k is of the type (14), (18) provided that $u_{k-1}(x,y)$ is a function of class C^1 satisfying $|H^{-1}u_{k-1}|\leqq c$ on (4), where $0<\alpha=\alpha_{k-1}\leqq a$. Under this proviso, the system (25)-(25 bis) has a unique C^1-solution $u_k=u_k(x,y)$ on (4), with $\alpha=\alpha_{k-1}$, since F satisfies (8). The analogue of (17) shows that u_k satisfies the inequality

$$|u_k|\leqq\alpha\max|(H_x-FH_y)H^{-1}u_{k-1}+Hg(x,y,H^{-1}u_{k-1})|,$$

where max refers to (x,y) on (4). Since $|H^{-1}u_{k-1}|\leqq c$, this can be written as $|u_k|\leqq(MC+M)\alpha$. Hence, if α is chosen so that

$$(MC+M)\alpha\leqq d, \tag{26}$$

where

$$d=c/n\max|H^{-1}|\leqq c/M, \tag{27}$$

then $z_k=H^{-1}u_k$ satisfies $|z_k|\leqq c$. Thus there exists a positive $\alpha(\leqq a)$, independent of k, such that (25)-(25 bis) has on (4) a C^1-solution $u_k(x,y)$ satisfying

$$|u_k|\leqq(MC+M)x\leqq d \qquad (k=1,2,\cdots). \tag{28}$$

If $k\geqq 2$, then, according to (25)-(25 bis), the difference $v_k=u_k-u_{k-1}$ is a solution of the initial value problem

$$v_{kx}=Fv_{ky}+G, \qquad v_k(0,y)=0, \tag{29}$$

where

$$G=(H_x-FH_y)H^{-1}v_{k-1}+H\{g(x,y,H^{-1}u_{k-1})-g(x,y,H^{-1}u_{k-2})\}.$$

Hence $|G|\leqq(MC+M)|v_{k-1}|$. A comparison of (29) and (14) shows that, in view of the analogue of (17),

$$|v_k|\leqq(MC+M)\int_0^x|v_{k-1}|\,dx.$$

Since $u_0\equiv 0$ and $v_1=u_1-u_0\equiv u_1$, the inequality (28) implies that $|v_1|\leqq(MC+M)x$. A simple induction then gives $|v_k|\leqq(MC+M)^kx^k/k!$ on (4) for $k=1,2,\cdots$. Consequently, $u_0,u_1,\cdots$ converge uniformly on (4) to a continuous function, say $u=u(x,y)$, which satisfies

$$|u| \leqq (MC+M)x. \tag{30}$$

It is clear from (25) that the existence proof for the semi-linear case will be complete if it is verified that the sequences of partial derivatives $u_{0x}, u_{1x}, \cdots$ and $u_{0y}, u_{1y}, \cdots$ converge uniformly on (4). In view of the uniform convergence of $u_0, u_1, \cdots$, it is therefore sufficient to show that the sequences of partial derivatives are uniformly bounded and equicontinuous.

In order to prove their uniform boundedness, note that

$$|u_{kx}| \leqq K|u_{ky}| + MC + M, \tag{31}$$

by (25), (27) and (28). Consequently,

$$\mu_{k1}(x) \leqq K\mu_{k2}(x) + MC + M, \tag{32}$$

if $\mu_{k1}(\beta)$, $\mu_{k2}(\beta)$ denote the maximum of $|u_{kx}|$, $|u_{ky}|$, respectively, on that part of (4) on which $0 \leqq x \leqq \beta (\leqq a)$. An upper estimate of u_{ky} can be obtained from an analogue of (22), where $z = u_k$, $U = H^{-1}u_{k-1}$ and $V = Hg(x, y, H^{-1}u_{k-1})$. The exponential factor in (22) is majorized by $e^{(MC+M)a}$, which does not exceed an M, by (26) and (27). The non-integrated term H_yU of (22) is majorized by $(MC+M)|u_{k-1}| \leqq (MC+M)^2x$ in view of (28). The term H_xU_y under the integral sign has an absolute value which is not greater than $(MC+M)(M\mu_{k-1,2} + MC + M)$, which can be written as $(MC+M)^2 + (MC+M)\mu_{k-1,2}$. Similarly,

$$|H_yU_x| \leqq (MC+M)^2 + (MC+M)\mu_{k-1,1}.$$

Finally, $|V_y| \leqq MC + M + M\mu_{k-1,2}$. Consequently,

$$\mu_{k2}(x) \leqq (MC+M)^2x + (MC+M)\int_0^x (\mu_{k-1,1} + \mu_{k-1,2})\,dt. \tag{33}$$

The relation (32), with k replaced by $k-1$, and (33) imply that

$$\mu_{k2}(x) \leqq (MC+M)^2x + (MC+M)^2\int_0^x \mu_{k-1,2}\,dt.$$

Since $\mu_{0,2}(x) \equiv 0$, a simple induction gives

$$\mu_{k2}(x) \leqq \sum_{j=1}^{k} (MC+M)^{2j}x^j/j! \leqq -1 + \exp(MC+M)^2x. \tag{34}$$

The definitions of μ_{k1} and μ_{k2} and the inequalities (31) and (34) show that the sequences $u_{0x}, u_{1x}, \cdots$ and $u_{0y}, u_{1y}, \cdots$ are uniformly bounded on (4).

There remains the problem of proving the equicontinuity of these sequences.

In view of the uniform convergence of the sequence $u_0, u_1, \cdots$, there exists, for small positive δ, a positive monotone function $w(\delta)$, which tends to 0 as $\delta \to +0$ and exceeds the moduli of continuity of $H, H_x, H_y, F, g, g_x, g_y, u_0, u_1, \cdots$ and of those combinations of the latter functions which will occur below. (By the modulus of continuity of a vector or matrix is meant the maximum of the moduli of continuity of its components or elements.) Hence (25) shows that

$$w_{k1}(\delta, x) \leqq K w_{k2}(\delta, x) + A w(\delta), \tag{35}$$

where A is a constant and $w_{k1}(\delta, \beta)$, $w_{k2}(\delta, \beta)$ are the moduli of continuity of u_{kx}, u_{ky} on that portion of (4) on which $0 \leqq x \leqq \beta (\leqq \alpha)$.

Straightforward estimates, analogous to those leading to (33), give

$$w_{k2}(\delta, x) \leqq A w(\delta) + A \int_0^x (w_{k-1,1} + w_{k-1,2})\, dt \tag{36}$$

with some constant A (which is a bound for certain combinations of H, F, g and their partial derivatives). It follows from the last two formula lines that

$$w_{k2}(\delta, x) \leqq A w(\delta) + A \int_0^x w_{k-1,2}(\delta, t)\, dt$$

(if A is an appropriate constant, independent of k, x and δ). Hence $w_{02} \equiv 0$ implies that

$$w_{k2}(\delta, x) \leqq A w(\delta) \sum_{j=0}^{k} A^j x^j / j! \leqq A w(\delta) e^{Ax}. \tag{37}$$

The definitions of w_{k1}, w_{k2} and the inequalities (35), (37) prove the equicontinuity of $u_{0x}, u_{1x}, \cdots$ and $u_{0y}, u_{1y}, \cdots$ on (4).

This completes the proof of the semi-linear case of (VI).

Remark 1. The number α determining the region (4) on which the solution $u(x, y)$ was proved to exist is subject only to the inequalities $0 < \alpha \leqq a$ and (26).

Remark 2. The inequality (34) shows that the partial derivative u_y of the solution $u = u(x, y)$ of (25)-(25 bis), the existence of which was just proved, satisfies the inequality

$$| u_y(x, y) | \leqq -1 + \exp(MC + M)^2 \tag{38}$$

on (4). Hence, by (25) and (30),

$$| u_x(x, y) | \leqq K\{-1 + \exp(MC + M^2)\alpha\} + (MC + M)^2 \alpha + M. \tag{39}$$

12. Proof of (VI). The existence proof just completed shows that if $z_0(x,y)=0$, and if $z_{k-1}(x,y)$ is a function of class C^1 satisfying $|z_{k-1}|\leqq c$ on (4), where $a=a_{k-1}$, $0<a_{k-1}\leqq a$, then the initial value problem

$$(40)\qquad z_{kx}=f(x,y,z_{k-1}(x,y))z_{ky}+g(x,y,z_k),\qquad z_k(0,y)=0$$

has a C^1-solution $z_k(x,y)$ with the same properties on (4), with some $a=a_k$ and $0<a_k\leqq a_{k-1}$, for $k=1,2,\cdots$. In order to make the inequalities of the last section more readily available, transform (40) by the change of variables (23). Thus (40) becomes

$$(41)\qquad u_{kx}=F_ku_{ky}+(H_{kx}-F_kH_{ky})H_k^{-1}u_k+H_kg(x,y,H_k^{-1}u_k),$$

$$(41\text{ bis})\qquad u_k(0,y)=0,$$

where $u_k=H_kz_k$, $F_k=F(x,y,z_{k-1}(x,y))$, and $H_k=H(z,y,z_{k-1}(x,y))$; cf. the transformation from (1) to (25). Remark 1 at the end of the last section shows that $a=a_k$ is subject only to the restrictions $0<a_k\leqq a_{k-1}$ and (26), where $C=C_{k-1}$ is a bound of the absolute value of the partial derivatives of the elements of the matrices H_k, F_k, and M is a fixed upper bound of certain combinations of f, g, H, F and their partial derivatives (in which x, y, z are considered as independent variables on (2_1)-(2_2)).

In view of the Remark made at the beginning of Section **11**,

$$(42)\qquad C_k\leqq M+MB_k,$$

where B_k is any upper bound for $|u_{kx}|$, $|u_{ky}|$ on (4), with $a=a_k$. It will be shown by induction that there exist positive numbers a, B (independent of k) with the properties that

$$(43_k)\qquad a_k\geqq a>0,$$

and

$$(44_k)\qquad |u_{kx}|\leqq B\text{ and }|u_{ky}|\leqq B$$

for $k=0,1,\cdots$ on (4).

Let the bound M be fixed so large as to satisfy all of the inequalities involving M in Section **11**, and put $B=2M$. Let $a(\leqq a_0)$ be a positive number satisfying the inequality

$$(45)\qquad (MC^*+M)a\leqq d,\text{ where }C^*=M+MB,\qquad d=c/M$$

(cf. (27)), as well as the inequalities

$$(46)\qquad -1+\exp(MC^*+M)^2a\leqq B,$$

$$(47)\qquad K\{-1\exp(MC^*+M)^2a\}+(MC^*+M)^2a+M\leqq B.$$

The inequalities (43_k), (44_k) are trivially satisfied when $k=0$. Suppose that they hold for some fixed $k \geqq 0$. In view of (42) and the fact that B_k can be taken to be B, it is seen that $C_k \leqq C^*$. It follows from (45) and the sentence following (41 bis) above that a_{k+1} can be chosen so as to satisfy (43_{k+1}). Finally, (44_{k+1}) is a consequence of (38), (39) in Remark 2 at the end of Section **11**, and of (46) and (47). This completes the induction.

Thus the functions $z_0(x,y), z_1(x,y), \cdots$ are defined (and of class C^1) on (4), where $a(>0)$ is independent of k. In addition, $|z_k| \leqq c$ for $k=0,1,\cdots$, and the sequences $z_{0x}, z_{1x}, \cdots$ and $z_{0y}, z_{1y}, \cdots$ are uniformly bounded on (4).

It will be proved that $z_0, z_1, \cdots$ converges uniformly on (4). To this end, let $\zeta_k = z_k - z_{k-1}$. Then ζ_k satisfies

$$\zeta_{kx} = f_k \zeta_{ky} + (\Delta_k f)\zeta_{k-1,y} + \Delta_k g, \qquad \zeta_k(0,y) = 0,$$

where

$$\Delta_k f = f_k - f_{k-1},\ \Delta_k g = g_k - g_{k-1}, \text{ and } f_k = f(x,y,z_{k-1}(x,y)),$$
$$g_k = g(x,y,z_{k-1}(x,y)).$$

Let $v_k = H_k \zeta_k$, where $H_k = H(x,y,z_{k-1}(x,y))$. Then v_k satisfies the partial differential equation

$$v_{kx} = F_k v_{ky} + (H_{kx} - F_k H_{ky}) H_k^{-1} v_k + H_k\{(\Delta_k f)\zeta_{k-1,y} + \Delta_k g\};$$

cf. the transformation from (1) to (25). Since $|\Delta_k f| \leqq \text{Const.}\, |v_{k-1}|$ and $|\Delta_k g| \leqq \text{Const.}\, |v_{k-1}|$, it follows from an analogue of (17) that

$$|v_k| \leqq a \text{ const.} (|v_k| + |v_{k-1}|),$$

where const. (independent of k) is a bound for combinations of H_k and the partial derivatives of H_k, F_k, g. If $r = a \text{ const.} < 1$ then

$$|v_k| \leqq r\, |v_{k-1}|/(1-r).$$

Hence, by induction,

$$|v_k| \leqq A[r/(1-r)]^{k-1},$$

where A is a bound of $|v_1|$. Consequently, if a is so small that $r/(1-r) < 1$, then $v_1 + v_2 + \cdots$ is absolutely and uniformly convergent on (4). Hence the same is true of $\zeta_1 + \zeta_2 + \cdots$. In view of the definition of ζ_k, this proves that the sequence $z_0, z_1, \cdots$ has a uniform limit, say $z(x,y)$, on (4).

It is clear that the proof of the existence statement in (VI) will be complete if it is shown that the sequences $z_{0x}, z_{1x}, \cdots$ and $z_{0y}, z_{1y}, \cdots$ are uniformly convergent on (4). Since these sequences are uniformly bounded and $z_0, z_1, \cdots$ is uniformly convergent, it is sufficient to prove that $z_{0x}, z_{1x}, \cdots$ and $z_{0y}, z_{1y}, \cdots$ are equicontinuous or, what is the same thing, that $u_{0x}, u_{1x}, \cdots$

and $u_{0y}, u_{1y}, \cdots$ are equicontinuous. But the proof of this fact is analogous to the corresponding proof in Section **11** (if $a > 0$ is sufficiently small) and will therefore be omitted.

The proof of the uniqueness statement in (*) follows the standard arguments which are applicable when successive approximations converge.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] L. Bianchi, *Lezioni di geometria differenziale*, Bologna, 3rd ed. (1922-23).

[2] M. Cinquini-Cibrario, "Sopra la teoria delle caratteristiche per i sistemi di equazioni quasi-lineari alle derivate parziali del primo ordine," *Annali della Scuola Normale Superiore de Pisa*, ser. 3, vol. 3 (1950), pp. 161-197.

[3] G. Darboux, *Leçons sur la théorie générale des surfaces*, vol. 3, Paris (1894).

[4] K. O. Friedrichs, "Nonlinear hyperbolic differential equations for functions of two independent variables," *American Journal of Mathematics*, vol. 70 (1948), pp. 555-589.

[5] E. Goursat, *Leçons sur l'intégration des équations aux derivées partielles du second ordre*, Paris (1896).

[6] J. Hadamard, *Le problème de Cauchy*, Paris (1932).

[7] P. Hartman and A. Wintner, "On the embedding problem in differential geometry," *American Journal of Mathematics*, vol. 72 (1950), pp. 553-563.

[8] ——— and A. Wintner, "On the fundamental equations of differential geometry," *ibid.*, pp. 757-774.

[9] ——— and A. Wintner, "'Gaussian curvature and local embedding," *ibid.*, vol. 73 (1951), pp. 876-884.

[10] ——— and A. Wintner, "On the embedding of hyperbolic line elements; a correction," *ibid.*, vol. 74 (1952), p. 264.

[11] E. Holmgren, "Sur la théorie des équations integrales linéaires," *Arkiv för Matematik, Astronomi och Fysik*, vol. 3 (1906), no. 1.

[12] ———, "Sur les systèmes linéaires aux derivées partielles du premier ordre," *ibid.*, vol. 6 (1911), no. 2.

[13] E. R. van Kampen, "Remarks on ordinary differential equations," *American Journal of Mathematics*, vol. 59 (1937), pp. 144-152.

[14] H. Lewy, "Ueber das Anfangswertproblem einer hyperbolischen nichtlinearen partiellen Differentialgleichung zweiter Ordnung mit zwei unabhängigen Veränderlichen," *Mathematische Annalen*, vol. 98 (1928), pp. 179-191.

[15] P. Montel, "Sur les suites infinies de fonctions," *Annales de L'École Normale Supérieure*, ser. 3, vol. 24 (1907), pp. 233-334.

[16] O. Perron, "Ueber Existenz und Nichtexistenz von Integralen partieller Differentialgleichungssysteme im reellen Gebiet," *Mathematische Zeitschrift*, vol. 27 (1928), pp. 549-564.

ON PICARD VARIETIES.*

By André Weil.

In an important recent paper [7], J. Igusa has defined two abelian varieties attached to a non-singular algebraic subvariety V of a complex projective space, and proved some basic results concerning these varieties and their relationship to each other. In the language of the Italian school, these varieties are those attached respectively to the periods of the simple integrals of the first kind on V, and to the continuous systems on V. It is my purpose here to present a somewhat different approach to the same problem, leading to a proof of Igusa's theorems under more general conditions, and to further results in the same direction. I shall avail myself freely of some of Igusa's ideas, but shall not otherwise make any use of his paper.

§ I. Igusa's duality theorems.

1. A vector space A of dimension q over the field $\boldsymbol{C}$ of complex numbers is at the same time a vector-space, which we shall denote by A_0, of dimension $2q$ over the field $\boldsymbol{R}$ of real numbers; the scalar multiplication by i in A is an automorphism J of A_0, satisfying $J^2 = -I$, where I is the identical automorphism of A_0. Conversely, if A_0 is a vector-space of dimension $2q$ over $\boldsymbol{R}$, and J is a linear mapping of A_0 into itself, satisfying $J^2 = -I$, we can define on A_0 a vector-space structure over $\boldsymbol{C}$ by putting $(\alpha + i\beta)x = \alpha x + \beta Jx$ for any $\alpha \varepsilon \boldsymbol{R}$, $\beta \varepsilon \boldsymbol{R}$; with this additional structure, A_0 will be denoted by A. Linear forms on A_0 will be called real-linear; linear forms on A will be called complex-linear. The complex-linear forms on A are those complex-valued real-linear forms $L(x)$ on A_0 which satisfy $L(Jx) = iL(x)$; if $M(x)$ is any real-valued real-linear form on A_0, then there is one and only one complex-linear form on A with the real part $M(x)$, viz. $L(x) = M(x) - iM(Jx)$.

A and A_0 being as above, let A'_0 be the dual space of A_0; denote by $\langle x', x\rangle$ the canonical bilinear form on $A'_0 \times A_0$. Put $J' = {}^tJ^{-1} = -{}^tJ$; as we have $J'^2 = -I'$, where I' is the identical automorphism of A'_0, J' can be used to make of A'_0 a complex vector-space A'. Put

$$B(x', x) = \tfrac{1}{2}[\langle x', x\rangle - i\langle x', Jx\rangle] = \tfrac{1}{2}[\langle x', x\rangle + i\langle J'x', x\rangle];$$

* Received December 21, 1951.

for each $a' \varepsilon A'$, $B(a', x)$ is a complex-linear form on A, and every such form can be so written in one and only one way; similarly, for each $a \varepsilon A$, $\overline{B(x', a)}$ is a complex-linear form on A', and every such form can be so written in one and only one way. We shall say that the sesquilinear form $B(x', x)$ defines A' as the antidual space of A.

2. By a *complex torus* Θ, of complex dimension q, we shall understand a torus of real dimension $2q$, with a translation-invariant complex-analytic structure. The universal covering group of Θ is then a vector-space of dimension $2q$ over $\boldsymbol{R}$ with a translation-invariant complex-analytic structure, or, what is the same thing, a vector-space A of dimension q over $\boldsymbol{C}$. In other words, a complex torus Θ is the quotient of a vector-space A of dimension q over $\boldsymbol{C}$ by a discrete subgroup Δ of rank $2q$ of A. If $\Theta_1 = A_1/\Delta_1$, $\Theta_2 = A_2/\Delta_2$ are two such complex toruses, a complex-analytic homomorphism of Θ_1 into Θ_2 will be given by a complex-linear mapping L of A_1 into A_2 such that $L(\Delta_1) \subset \Delta_2$. If Θ_1, Θ_2 have the same dimension, and if there exists an *invertible* complex-linear mapping L of A_1 into A_2 such that $L(\Delta_1) \subset \Delta_2$, then we say that Θ_1, Θ_2 are *isogenous*; as L must then map Δ_1 isomorphically onto a subgroup of finite index of Δ_2, there will then be an integer $n \neq 0$ such that nL^{-1} maps Δ_2 into Δ_1; it follows that isogeneity is an equivalence relation between complex toruses (to this relation, an unfortunate tradition had attached the word "isomorphism," which of course is here restored to its proper meaning).

3. Let Θ be a complex torus, of complex dimension q; let A be the universal covering group of Θ, considered as above as a vector-space of dimension q over $\boldsymbol{C}$. Let A_0 be the underlying real vector-space of A, of dimension $2q$ over $\boldsymbol{R}$, and let J be the automorphism of A_0 defined by the scalar multiplication by i on A. Let Δ be the kernel of the canonical homomorphism of A onto Θ, so that $\Theta = A/\Delta$. Let A' be the antidual space to A, as defined in no. 1, so that its underlying real vector-space A'_0 is the dual of A_0, and that A' is determined over A'_0 by the automorphism $J' = {}^tJ^{-1} = -{}^tJ$. Let Δ' be the subgroup of A'_0 associated with the subgroup Δ of A_0, i. e. consisting of all the elements r' of A'_0 such that $\langle r', r\rangle \equiv 0 \bmod 1$ for all $r \varepsilon \Delta$. Then Δ' is a discrete subgroup of rank $2q$ of A'; and $\Theta' = A'/\Delta'$ is a complex torus, which will be called *dual* to Θ. It is clear that Θ is then dual to Θ'

Let $\Theta_1 = A_1/\Delta_1$, $\Theta_2 = A_2/\Delta_2$ be two complex toruses, and Θ'_1, Θ'_2 their duals. Consider a complex-analytic homomorphism of Θ_1 into Θ_2, given as above by a complex-linear mapping L of A_1 into A_2, such that $L(\Delta_1) \subset \Delta_2$.

But, if A_1, A_2 are defined by the underlying real vector-spaces and the automorphisms J_1, J_2 of these corresponding to the scalar multiplication by i, a complex-linear mapping of A_1 into A_2 is the same as a real-linear mapping L_0 satisfying $L_0J_1 = J_2L_0$; as this implies ${}^tL_0(-{}^tJ_2) = (-{}^tJ_1){}^tL_0$, it follows that ${}^tL_0$ is a complex-linear mapping of A'_2 into A'_1; and one verifies at once that ${}^tL_0(\Delta'_2) \subset \Delta'_1$. Therefore ${}^tL_0$ defines a complex-analytic homomorphism of Θ'_2 into Θ'_1.

4. With the same notations as above, let $\Theta = A/\Delta$ be a complex torus of complex dimension q. Let Γ be the group of all representations of the additive group Δ into the multiplicative group $\boldsymbol{C}^*$ of complex numbers $\neq 0$; let Γ_0 be the subgroup of the elements μ of Γ such that $|\mu(r)| = 1$ for all $r \varepsilon \Delta$, i. e. the character-group of Δ. As Δ is a free abelian group of rank $2q$, Γ_0 is a torus of real dimension $2q$, and Γ is isomorphic to $(\boldsymbol{C}^*)^{2q}$. Elements of Γ will be called *multiplicator-sets.* By a multiplicative function belonging to Θ, with the multiplicator-set μ, we shall understand a complex-valued function ϕ on A, not everywhere 0 or ∞, such that $\phi(x+r) = \phi(x)\mu(r)$ for all $x \varepsilon A$, $r \varepsilon \Delta$. A multiplicative function which is an entire (i. e., everywhere holomorphic) function of the complex coordinates on A, and is nowhere 0, must be of the form[1] $e[L(x)+c]$, where c is a constant and $L(x)$ a complex-linear form on A; the multiplicator-set $e[L(r)]$ of any such function will be called *trivial*; we shall denote by γ the group of all trivial multiplicator-sets.

If $\mu \varepsilon \Gamma$, $\log|\mu(r)|$ is a representation of the additive group Δ into the additive group $\boldsymbol{R}$; this can be extended in one and only one way to a real-valued real-linear form $M(x)$ on A_0; let $L(x) = M(x) - iM(Jx)$ be the complex-linear form on A with the real part $M(x)$. Then $\nu(r) = e^{L(r)}$ is a trivial multiplicator-set, satisfying $|\nu(r)| = |\mu(r)|$ for all $r \varepsilon \Delta$, and it is the only such set. Hence every coset of γ in Γ has one and only one element in Γ_0, i. e. we have $\Gamma = \gamma \times \Gamma_0$; and Γ/γ may be canonically identified with Γ_0.

Let A', A'_0, Δ' be as in no. 3. Each element of Γ_0 can be written as $e(\langle a', r\rangle)$, with $a' \varepsilon A'$; this defines therefore a homomorphism of A' onto Γ_0, the kernel of which is Δ', hence an isomorphism between Γ_0 and A'/Δ'. Thus we have canonical isomorphisms between Γ/γ, Γ_0, and the underlying real torus of $\Theta' = A'/\Delta'$; we can use these to transport to Γ/γ and Γ_0 the complex torus structure of Θ', and to identify these three toruses with one another.

---

[1] Here and in the rest of this paper we put $e(t) = e^{2\pi i t}$.

5. On Γ, one can define a complex-analytic structure in an obvious manner, viz. by the condition that, for each $r \varepsilon \Delta$, $\mu(r)$, as a function of $\mu \varepsilon \Gamma$, should be a holomorphic function on Γ. We now prove that the canonical homomorphism of Γ onto Γ/γ is complex-analytic, when the complex structure on Γ/γ is defined as in no. 4, and that this complex structure on Γ/γ is the only one for which this is so. In fact, let B' be the vector-space of dimension $2q$ over $\boldsymbol{C}$, derived from A'_0 by extending the field of scalars of A'_0 from $\boldsymbol{R}$ to $\boldsymbol{C}$; every element of B' can be written as $z' = x' + iy'$, with $x' \varepsilon A'_0$, $y' \varepsilon A'_0$; if we write $\mu_{z'}(r) = e(\langle x', r\rangle + i\langle y', r\rangle)$, the mapping $z' \to \mu_{z'}$ is a complex-analytic homomorphism of B' onto Γ. Now the element ν of γ which satisfies $|\nu(r)| = |\mu_{z'}(r)|$ for all $r \varepsilon \Delta$ is $\nu(r) = e(\langle y', Jr\rangle + i\langle y', r\rangle)$; hence the element of Γ_0 which belongs to the same coset of γ as $\mu_{z'}$ is $e(\langle x', r\rangle - \langle y', Jr\rangle) = e(\langle x' + J'y', r\rangle)$. Now denote that coset by $f(z')$; f is a homomorphism of B' onto Γ/γ. Call ω the canonical homomorphism of A' onto $\Theta' = A'/\Delta'$; and, for $z' = x' + iy'$, put $g(z') = x' + J'y'$; Θ' and Γ/γ being identified as in no. 4, what we have just shown amounts to $f = \omega \circ g$. As we have $g(iz') = -y' + J'x' = J'(x' + J'y')$, g is a complex-analytic homomorphism; so is ω; the same, therefore, is true of f. Conversely, if J'_1 is an automorphism of A'_0, defining on A'_0 a complex structure such that f, hence also g, are complex-analytic mappings, then we must have $g(iz') = J'_1 g(z')$, hence $J'_1 = J'$.

6. From the theory of abelian varieties and of theta-functions ([8], [9]; cf. [11]), we borrow the following facts. In order that there should be q algebraically independent meromorphic functions on the complex torus $\Theta = A/\Delta$, it is necessary and sufficient that there should exist a real-valued alternating bilinear form $E(x, y)$ on $A_0 \times A_0$, such that $E(r, s) \equiv 0 \bmod 1$ whenever $r \varepsilon \Delta, s \varepsilon \Delta$, and that the bilinear form $F(x, y) = E(Jx, y)$ is symmetric and positive-definite (i. e. satisfies $F(x, x) > 0$ for all $x \neq 0$). Such a form E will be called a *Riemann form* for Θ; if we write it as $E(x, y) = \langle Ex, y\rangle$, where E is a linear mapping of A_0 into A'_0, the condition that it should be alternating is expressed by ${}^tE = -E$; this being assumed, we have $F(x, y) = \langle EJx, y\rangle = -\langle Ey, Jx\rangle$, and the condition that F should be symmetric is expressed by $EJ = {}^t(EJ) = J'E$, hence it is equivalent to the condition that E should be a complex-linear mapping of A into A'. The condition $E(r, s) \equiv 0 \bmod 1$ for $r \varepsilon \Delta$, $s \varepsilon \Delta$, is equivalent to $E(\Delta) \subset \Delta'$. Hence E defines a complex-analytic homomorphism of Θ into its dual Θ'. Finally, if F is positive-definite, we must have $Ex \neq 0$ for $x \neq 0$, hence E is invertible. Therefore the existence of a Riemann form implies that Θ and Θ' are isogenous; if moreover E maps Δ onto Δ', they will be isomorphic, but they need not be so otherwise.

A Riemann form E being given, one can define (by the explicit con-

struction of so-called theta-series) theta-functions belonging to Θ and E, i. e. entire functions ϑ of the complex coordinates on A, satisfying periodicity conditions which, suitably normalized, may be assumed to be

$$\vartheta(x+r)=\vartheta(x)e[-\tfrac{1}{2}E(r,x)+\tfrac{i}{2}E(r,Jx)-\tfrac{1}{4}E^0(r,r)+\tfrac{i}{4}E(r,Jr)]$$

for all $x \varepsilon A, r \varepsilon \Delta$, where $E^0(x,y)$ is a symmetric bilinear form on $A_0 \times A_0$ such that $E^0(r,s) \equiv E(r,s) \bmod 2$ for all $r \varepsilon \Delta, s \varepsilon \Delta$. Then, for any $a \varepsilon A$, the function

$$\vartheta(x+a)\vartheta(x)^{-1}e[\tfrac{1}{2}E(a,x)-\tfrac{i}{2}E(a,Jx)]$$

is a multiplicative meromorphic function of the complex coordinates on A, with the multiplicator-set $e[E(a,r)]=e(\langle Ea,r\rangle)$. As E maps A_0 onto A'_0, this, in conjunction with the results of no. 4, shows that there exist meromorphic multiplicative functions with arbitrarily given multiplicator-sets.

Now consider a maximal set of linearly independent theta-functions $\vartheta_j(x)$ belonging to Θ and the Riemann form NE, N being an integer >0. These must be in finite number; and, if they are taken as homogeneous coordinates of a point $P(x)$ in a suitable projective space, we have $P(x+r)=P(x)$ for $r \varepsilon \Delta$, so that the mapping $x \to P(x)$ induces a complex-analytic mapping Φ of Θ into that projective space. By a theorem of Lefschetz [8], as soon as N is large enough, Φ is a one-to-one mapping of Θ onto an algebraic variety $\Phi(\Theta)$, and its Jacobian matrix has everywhere the rank q, so that $\Phi(\Theta)$ is without multiple points; Φ can then be used to identify Θ with $\Phi(\Theta)$. For that reason, we shall say that a complex torus Θ is an *abelian variety* whenever there exists a Riemann form for Θ.

If Θ is an abelian variety, so is every complex torus isogenous to Θ, hence in particular the dual torus Θ'. In fact, let $\Theta_1=A_1/\Delta_1$ be isogenous to Θ; then there is an invertible complex-linear mapping L of A_1 onto A, such that $L(\Delta_1)\subset\Delta$. Let $E(x,y)$ be a Riemann form for Θ; then the bilinear form $E(Lx_1,Ly_1)$ is alternating, and is $\equiv 0 \bmod 1$ for $x_1 \varepsilon \Delta_1$, $y_1 \varepsilon \Delta_1$; also, as $LJ_1=JL$, we have

$$E(LJ_1x_1,Ly_1)=E(JLx_1,Ly_1)=F(Lx_1,Ly_1),$$

where F is defined as before; this is symmetric, and >0 whenever $Lx_1=Ly_1 \neq 0$, hence, since L is invertible, whenever $x_1=y_1\neq 0$.

7. Now let V be a Kähler manifold, i. e. a connected compact complex-analytic manifold with a Kähler metric; I shall use the notations of a previous note of mine ([10][2]; cf. [3], Part II).

[2] In that note, one should read $\Omega=\frac{i}{2}\Sigma_\nu\omega_\nu\wedge\bar{\omega}_\nu$ instead of $\Omega=\Sigma_\nu\omega_\nu\wedge\bar{\omega}_\nu$ on p. 110, line 7 from bottom.

If G is the fundamental group of V, and G' the commutator-group of G, $H = G/G'$ is the one-dimensional homology group of V with integral coefficients; it is the direct product of its torsion group T, i. e. of the finite group of its elements of finite order, and of a free abelian group, whose rank, by Hodge's theorems, is an even number $2q$. If we consider H as a module over the ring of integers, and extend[3] its ring of operators to $\boldsymbol{R}$, we obtain a vector-space A_0 of dimension $2q$ over $\boldsymbol{R}$, which is the one-dimensional homology group of V with real coefficients; we shall denote by λ the canonical homomorphism of H into A_0; the kernel of λ is T, and the image $\Delta = \lambda(H)$ of H by λ is a discrete subgroup of rank $2q$ of A_0.

Let $\hat{V}$ be the covering manifold of V belonging to the subgroup G' of G; every element σ of H determines an automorphism of $\hat{V}$, transforming each point $\hat{M}$ of $\hat{V}$ into a point $\sigma\hat{M}$ lying over the same point M of V as $\hat{M}$. If θ is any closed differential form of degree 1 on V, and $\hat{P}$, $\hat{M}$ are two points on $\hat{V}$, the integral $\int_{\hat{P}}^{\hat{M}} \theta$ (taken along any differentiable path from $\hat{P}$ to $\hat{M}$ on $\hat{V}$) defines, when $\hat{P}$ is fixed, a function $f(\hat{M})$ such that $df = \theta$; we have simplified notations here by writing θ also for the inverse image of θ on $\hat{V}$. For any $\sigma \in H$, $p_\sigma = f(\sigma\hat{M}) - f(\hat{M})$ is a constant, which we can also write as $p_\sigma = \int_\sigma \theta$, the integral being taken along any closed differentiable path of class σ on V. Then $\sigma \to p_\sigma$ is a representation of H into the additive group $\boldsymbol{R}$; this must be 0 on T, and it can be extended, in one and only one way, to a linear form on A_0; this form will be denoted by $p(\theta, x)$, so that we have $p_\sigma = \int_\sigma \theta = p[\theta, \lambda(\sigma)]$.

Now let B be the vector-space of real harmonic differential forms of degree 1 on V; then $p(\theta, x)$, as a function of $(\theta, x) \in B \times A_0$, is a bilinear form, which, by Hodge's fundamental theorem, determines a duality between B and A_0; hence we may use it to identify B with the dual space A'_0 of A_0, i. e. with the one-dimensional cohomology group of V with real coefficients; and we write, from now on, $\langle\theta, x\rangle$ instead of $p(\theta, x)$ when θ is harmonic, i. e. when $\theta \in A'_0$.

If θ is harmonic, and if $\hat{P}$, $\hat{M}$ are two points on $\hat{V}$, the integral $\int_{\hat{P}}^{\hat{M}} \theta$ is a real number depending linearly upon θ, which, when $\hat{P}$ is fixed, can be written as $\langle\theta, \hat{F}(\hat{M})\rangle$, with $\hat{F}(\hat{M}) \in A_0$. Then $\hat{F}$ is a mapping of $\hat{V}$ into A_0, satisfying $\langle\theta, \hat{F}(\sigma\hat{M}) - \hat{F}(\hat{M})\rangle = \int_\sigma \theta = \langle\theta, \lambda(\sigma)\rangle$, i. e. $\hat{F}(\sigma\hat{M}) = \hat{F}(\hat{M}) + \lambda(\sigma)$,

[3] Cf. N. Bourbaki, *Algèbre*, Chap. III, § 2, no. 1.

hence $\hat{F}(\sigma\hat{M}) \equiv \hat{F}(\hat{M}) \bmod \Delta$. Therefore $\hat{F}$ determines a mapping F of V into the torus A_0/Δ; $\hat{F}$ and F, or any mappings derived from these by the addition of constants, will be called the *canonical mappings* of $\hat{V}$ into A_0, and of V into A_0/Δ, respectively.

So far we have made no use of the complex structure of V. Now we introduce the operator C derived from this structure (cf. [10]); this transforms real differential forms into real differential forms of the same degree, harmonic forms into harmonic forms; hence it induces an automorphism on the space A'_0 of real harmonic forms of degree 1, whose transpose will be denoted by J. As we have $C(C\theta) = -\theta$ for any form of degree 1, we have $J^2 = -I$. Furthermore, if f is any holomorphic function in the neighborhood of a point of V, then we have $C(df) = i \cdot df$, $C(d\bar{f}) = -i \cdot d\bar{f}$; conversely, any closed form ζ of degree 1, satisfying $C\zeta = i\zeta$, is holomorphic, i. e. it is locally the differential of a holomorphic function. In particular, if θ is a real harmonic form, the form $\zeta = \theta - iC\theta$ is closed and satisfies $C\zeta = i\zeta$, hence it is everywhere holomorphic, i. e. it is a simple differential of the first kind; since, conversely, the real part of such a form is harmonic, this implies that the space A'_0 of real harmonic forms of degree 1 depends only upon the complex structure of V, and not upon the choice of a Kähler metric (provided one exists).

Now we use the automorphism J of A_0 in order to convert A_0 into a vector-space A of dimension q over $\mathbf{C}$, so that $\Theta = A/\Delta$ is then a complex torus. As any complex-linear form in A can be written as

$$L(x) = \langle\theta, x\rangle - i\langle\theta, Jx\rangle = \langle\theta, x\rangle - i\langle C\theta, x\rangle,$$

with some $\theta \in A'_0$, we have, for such a form, $L[\hat{F}(\hat{M})] = \int_{\hat{p}}^{\hat{M}} \zeta$ with $\zeta = \theta - iC\theta$, which shows that $L[\hat{F}(\hat{M})]$ is then a holomorphic function on $\hat{V}$. Therefore *$\hat{F}$ and F are complex-analytic mappings* of $\hat{V}$ into A, and of V into Θ, respectively.

8. Let Ω be the differential form of degree 2 on V canonically associated with the given Kähler metric (cf. [10]). If all the periods of Ω are integers (i. e., if Ω is homologous to an integral cocycle), we shall say that V satisfies the Hodge condition, or that it is a Hodge manifold. A recent theorem of Hodge [6] states that, when that is so, the complex torus Θ associated with V is an abelian variety. In our notation, Hodge's proof for this is as follows. Let $*$ be the usual operator, transforming forms of degree d into forms of degree $2n - d$ if n is the complex dimension of V (cf. [10]).

If θ is any differential form of degree 1, one verifies easily that $(C\theta)\Omega^{n-1} = -(n-1)! * \theta$, multiplication being of course the exterior multiplication of differential forms. Put now:

$$E'(\theta, \eta) = -\int_V \theta\eta\Omega^{n-1};$$

this is an alternating bilinear form on $A'_0 \times A'_0$. Call Δ' the subgroup of A'_0 associated with the subgroup Δ of A_0; it consists of those harmonic forms, all of whose periods are integers. As the periods of Ω are integers, one of de Rham's theorems shows that $E'(\theta, \eta) \equiv 0 \bmod 1$ for $\theta \in \Delta', \eta \in \Delta'$. If J' is defined as usual by $J' = -{}^tJ$, we have $J'\theta = -C\theta$ for $\theta \in A'_0$, hence

$$E'(J'\theta, \eta) = (n-1)! \int_V \eta \cdot (*\theta).$$

As this is symmetric and positive-definite, we see that E' is a Riemann form for Θ', so that Θ', hence also Θ, are abelian varieties. We shall call Θ' *the Picard variety*, and Θ *the dual Picard variety, of* V.

9. Henceforward, let V be a Hodge manifold. Let all notations have the same meaning as above; in particular, we still denote by λ the canonical homomorphism of H into A_0, and by $\hat{F}$, F the canonical mappings of $\hat{V}$ into A, and of V into Θ. Let ϵ be any representation of H into the multiplicative group $\boldsymbol{C}^*$; by a multiplicative function on V, with the multiplicator-set ϵ, we shall understand a complex-valued function ϕ on V, not everywhere 0 or ∞, such that $\phi(\sigma\hat{M}) = \phi(\hat{M})\epsilon(\sigma)$ for all $\hat{M} \in \hat{V}$, $\sigma \in H$. In particular, if ψ is a multiplicative function belonging to Θ, with the multiplicator-set μ, $\psi[\hat{F}(\hat{M})]$, provided it is not everywhere 0 or ∞, is a multiplicative function on $\hat{V}$ with the multiplicator-set $\epsilon = \mu \circ \lambda$; any multiplicator-set ϵ of that form will be called *special*, and any multiplicative function on $\hat{V}$ with such a multiplicator-set will be called *special*. A multiplicator-set ϵ is special if and only if it is equal to 1 on the torsion-group T. The mapping $\mu \to \mu \circ \lambda$ is an isomorphism of the group Γ of multiplicator-sets belonging to Θ onto the group Γ' of special multiplicator-sets for V; we shall denote by Γ'_0, γ' the images, by this isomorphism, of the subgroups Γ_0, γ of Γ as defined in no. 4; elements of γ' will be called the *trivial* multiplicator-sets for V.

If ψ is any holomorphic multiplicative function without zeros, belonging to Θ, and μ is its multiplicator-set, then $\mu \in \gamma, \mu \circ \lambda \in \gamma'$, and $\psi[\hat{F}(\hat{M})]$ is a holomorphic multiplicative function without zeros on $\hat{V}$, with the multiplicator-set $\mu \circ \lambda$. Conversely, let ϕ be a holomorphic multiplicative function

without zeros on $\hat{V}$; then $\zeta = \frac{1}{2\pi i} d(\log \phi)$ is a holomorphic differential on V, hence of the form $\zeta = \theta - iC\theta$, with $\theta \in A'_0$; then we have

$$\phi(\hat{M}) = e\left(\int_{\hat{P}}^{\hat{M}} \zeta + c\right) = \psi[\hat{F}(\hat{M})],$$

with

$$\psi(x) = e[L(x) + c], \qquad L(x) = \langle\theta, x\rangle - i\langle C\theta, x\rangle = \langle\theta, x\rangle - i\langle\theta, Jx\rangle,$$

where c is a constant; as L is a complex-linear form on A, this shows that the multiplicator-set of ϕ is trivial.

If ψ is any meromorphic multiplicative function belonging to Θ, with the multiplicator-set μ, one can choose a constant $a \in A$ so that $\psi[\hat{F}(\hat{M}) + a]$ is not everywhere 0 or ∞, and then this is a meromorphic multiplicative function on V with the multiplicator-set $\mu \circ \lambda$; in conjunction with the results of no. 6, this shows that there exist meromorphic multiplicative functions on $\hat{V}$ with an arbitrarily given special multiplicator-set.

10. One can define the *divisor* on V of any meromorphic multiplicative function on $\hat{V}$, or more generally of any meromorphic function ϕ on $\hat{V}$ such that, for every $\sigma \in H$, $\phi(\sigma\hat{M})/\phi(\hat{M})$ is a holomorphic function without zeros on $\hat{V}$; the definition is briefly as follows (cf. K. Kodaira, [3], Part II). If $z_1, \cdots, z_n$ are local complex coordinates in a neighborhood of a point P of V, and $\hat{P}$ is a point lying over P on $\hat{V}$, $z_1, \cdots, z_n$ can be used as local coordinates on $\hat{V}$ at $\hat{P}$, and ϕ can be expressed as a product $\phi = E(z)\prod_j \phi_j(z)^{m_j}$ of a unit-factor $E(z)$ (i. e. a holomorphic function, $\neq 0$ at P) and of powers of irreducible holomorphic functions $\phi_j(z)$ (irreducible, that is, in the ring of holomorphic functions at $\hat{P}$). Replacing $\hat{P}$ by another point $\sigma\hat{P}$ of $\hat{V}$ lying over P will merely affect $E(z)$, but not the $\phi_j(z)$ or the m_j. The divisor of ϕ on V will then be defined locally, in a neighborhood of P, as $\sum_j m_j W_j$, where W_j is the irreducible algebroid variety, of complex dimension $n-1$, defined by $\phi_j(z) = 0$, with the orientation determined by the condition $\Omega^{n-1} > 0$. Then the divisor (ϕ) of ϕ on V will be defined globally, by means of any suitable finite open covering of V; it will be of the form $\sum_\rho a_\rho Z_\rho$, where the a_ρ are integers, and the Z_ρ are irreducible compact analytic (i. e., everywhere algebroid) subvarieties of V, of complex dimension $n-1$, with the orientation defined above. More generally, any such expression will be called a divisor (more precisely, an analytic divisor) on V.

11. Now, let again ϕ be a meromorphic multiplicative function on $\hat{V}$, and put $\zeta = \frac{1}{2\pi i} d(\log \phi)$; it follows from the multiplicative property of ϕ that this is a meromorphic differential form on V (more accurately, it is the inverse image on $\hat{V}$ of such a form). By analytic continuation, one sees that $\phi(\hat{M}) = e(\int_{\hat{P}}^{\hat{M}} \zeta + c)$, where c is a constant; the integral is taken along any differentiable path, from the fixed point $\hat{P}$ on $\hat{V}$ to $\hat{M}$, which does not meet (ϕ); hence, for $\sigma \varepsilon H$, we have $\phi(\sigma\hat{M}) = \phi(M) e(\int_\sigma \zeta)$. If ϕ is factored as $\phi = E(z) \prod_j \phi_j(z)^{m_j}$ in terms of local coordinates in the neighborhood of a point of $\hat{V}$, then $\zeta - \sum_j m_j \frac{1}{2\pi i} d(\log \phi_j)$ is holomorphic in that neighborhood; from this, one deduces, by first considering the case of a singular 2-cell in such a neighborhood and of its boundary, that, if S is any 2-dimensional singular chain in V, whose boundary C is a differentiable singular chain and does not meet (ϕ), the intersection-number of S and (ϕ) is equal to $\int_C \zeta$; in the language of the theory of currents (cf. [3]), this shows that the cycle (ϕ) of dimension $2n-2$ on V is the differential of the current ζ. This implies, firstly, that (ϕ) is homologous to 0 with real coefficients, hence also (since it is an integral cycle [4]) with rational coefficients. In particular, a holomorphic multiplicative function ϕ on $\hat{V}$ cannot have zeros: for otherwise the integral of Ω^{n-1} on (ϕ) would be > 0, and, as this is a closed form, (ϕ) could not be homologous to 0. Also, if $\tau \varepsilon T$, the linking coefficient of τ and (ϕ) is equal to $\int_\tau \zeta \bmod 1$ (cf. Igusa [7]); for, if t is any differentiable singular cycle of class τ which does not meet (ϕ), that linking coefficient is by definition the intersection-number of (ϕ) and of any 2-dimensional singular chain with real coefficients, with the boundary t, reduced mod 1. As it is known, by Poincaré's duality theorem, that the linking coefficient defines the torsion groups of V, of dimensions 1 and $2n-2$, as a dual pair, it follows that (ϕ) is homologous to 0 with integral coefficients if and only if $\int_\tau \zeta \equiv 0 \bmod 1$ for all $\tau \varepsilon T$, i. e. if and only if the multiplicator-set of ϕ is special.

12. Let now Z be any analytic divisor on V. We can cover V with open sets U_h, so small that, in each U_h, Z can be written as $\sum_j m_j W_j$, where

[4] As to this and other "obvious" homological properties of the subvarieties of V and in particular of the divisors on V, they can best be justified by the definitions and results in N. Hamilton's forthcoming thesis [4].

W_j is the variety of zeros of an irreducible holomorphic function $\phi_j(z)$ in U_h, the z's being local complex coordinates in U_h. Define the differential form ζ_h in U_h by $\zeta_h = \sum_j m_j \frac{1}{2\pi i} d(\log \phi_j)$; then $\zeta_h - \zeta_k$ is holomorphic in $U_h \cap U_k$; hence, by the main theorem of [10], if Z is homologous to 0 with real coefficients (and only in that case), there will be a meromorphic differential form ζ on V, such that, in each U_h, $\zeta - \zeta_h$ is a closed holomorphic form. Analytic continuation and the monodromy principle then show that there exists a meromorphic multiplicative function ϕ on $\hat{V}$, such that, in each U_h, $\phi = E(z)\prod_j \phi_j(z)^{m_j}$, where $E(z)$ is a unit-factor in U_h, and that $\zeta = \frac{1}{2\pi i} d(\log \phi)$; then we have $(\phi) = Z$. By the results of no. 11, Z is homologous to 0 with integral coefficients if and only if the multiplicator-set of ϕ is special.

Now suppose that ϕ_1, ϕ are two meromorphic multiplicative functions such that $(\phi) = (\phi_1) = Z$; then ϕ_1/ϕ is a holomorphic multiplicative function without zeros, hence its multiplicator-set is trivial. Therefore, if $\mathfrak{G}_h$ is the group of the analytic divisors which are homologous to 0 with real coefficients, we have attached to each $Z \in \mathfrak{G}_h$ a coset $c(Z)$ of the group γ' of trivial multiplicator-sets in the group of all multiplicator-sets, consisting of the multiplicator-sets of all meromorphic multiplicative functions ϕ such that $(\phi) = Z$. This coset will consist of special multiplicator-sets, i. e. it will be in Γ'/γ', if and only if Z is in the subgroup $\mathfrak{G}_a$ of $\mathfrak{G}_h$ consisting of the divisors which are homologous to 0 with integral coefficients. We have $c(Z) = \gamma'$ if and only if Z is the divisor of a function ϕ with the multiplicator-set 1, i. e. of a meromorphic function on V; then we say that Z is linearly equivalent to 0; the group of such divisors will be denoted by $\mathfrak{G}_l$.

But Γ'/γ' can be canonically identified with Γ/γ, hence also with Γ_0 and Θ'. Therefore the mapping $Z \to c(Z)$, restricted to $\mathfrak{G}_a$, defines a canonical homomorphism of $\mathfrak{G}_a$ onto the Picard variety Θ' of V, with the kernel $\mathfrak{G}_l$; we may also say that we have defined a canonical isomorphism between $\mathfrak{G}_a/\mathfrak{G}_l$ and the character-group Γ_0 of Δ: this is Igusa's first duality theorem. Similarly, the mapping $Z \to c(Z)$ determines a canonical isomorphism between $\mathfrak{G}_h/\mathfrak{G}_l$ and a subgroup of finite index of the character-group of H; this subgroup will be the whole character-group of H whenever every character of H of finite order is the multiplicator-set of some meromorphic multiplicative function on $\hat{V}$, or, what amounts to the same in view of the above results, whenever every homology class of dimension $2n-2$ and of finite order on V (with integral coefficients) contains an analytic divisor. This is Igusa's second duality theorem. By a theorem of Lefschetz ([8]), V will have that property whenever it is a non-singular algebraic subvariety of a projective

space; indeed, it seems likely that, for such variety V, every covering manifold of V with a finite number of sheets is again an algebraic variety.

§ II. Construction of a system of representatives for the group of divisor-classes.

13. For the proof of the main result of this §, we shall need various lemmas on theta-functions. As explained in no. 10, if ϑ is a theta-function (other than 0) belonging to an abelian variety Θ, one can define the divisor (ϑ) of ϑ on Θ; this is a positive divisor, i. e. all its components have positive coefficients. If a point u of Θ lies on a component of (ϑ), we shall say that ϑ is 0 at u.

Periodicity conditions for theta-functions will always be understood to be in the normalized form given in no. 6. As this still depends upon the choice of a symmetric form E^0, when Θ and E are given, we shall further normalize this by choosing a set of generators $r_1, \cdots, r_{2q}$ for Δ, and taking $E^0(r_j, r_k) = E(r_j, r_k)$ for $1 \leqq j \leqq k \leqq 2q$; then E^0 depends linearly upon E. In applying this to a product $\Theta \times \Theta_1$ of two abelian varieties $\Theta = A/\Delta$, $\Theta_1 = A_1/\Theta_1$, it should be understood that, after having chosen the sets of generators for Δ, Δ_1, we take, as the generators for $\Delta \times \Delta_1$ those for Δ and those for Δ_1, in that order.

From this agreement it follows that, if E and E_1 are Riemann forms for Θ, and ϑ, ϑ_1 are theta-functions, belonging to Θ and E, and to Θ and E_1, respectively, then $\vartheta\vartheta_1$ is a theta-function belonging to $E + E_1$; and so is $\vartheta(x - E^{-1}E_1 a)\vartheta_1(x + a)$ for every $a \varepsilon A$, as one easily verifies. It is clear that one can choose a so that this is not 0 at a given point u of Θ, provided ϑ, ϑ_1 are not everywhere 0. In other words, if a Riemann form E is the sum of two Riemann forms, and if u is any point of Θ, there is a theta-function, belonging to Θ and E, which is not 0 at u.

14. If E is a Riemann form for Θ, the set of theta-functions belonging to Θ and E is a vector-space of finite dimension over $\boldsymbol{C}$. Let $(\vartheta_\mu)_{0 \leqq \mu \leqq M}$ be a basis for the space of theta-functions belonging to Θ and mE, where m is an integer > 0; by what has been said above, as soon as $m \geqq 2$, there is no point of Θ where all the ϑ_μ are 0; furthermore, by a theorem of Lefschetz which we have mentioned in no. 6, as soon as m has been taken large enough, the mutual ratios of the ϑ_μ cannot be the same at two points of A unless these points have the same image in Θ; let m be so chosen, once for all. Let now Z be an infinite subset of Θ; consider all the homogeneous polynomials

$\Phi(U_0,\cdots,U_M)$ in $M+1$ indeterminates, such that $\Phi(\vartheta_0,\cdots,\vartheta_M)$ is 0 on Z; they generate a homogenous ideal in the ring of all polynomials in $U_0,\cdots,U_M$, whose set of zeros, in the projective space of dimension M with the homogeneous coordinates $(u_0,\cdots,u_M)$, is infinite and therefore has at least one component of dimension >0. In the vector-space of all homogeneous polynomials Φ of degree r in $U_0,\cdots,U_M$, consider the subspace of those for which $\Phi(\vartheta_0,\cdots,\vartheta_M)$ is 0 on Z; let $\delta(r)$ be its codimension, i. e. the difference between the dimensions of the space and of the subspace. By Hilbert's theorem on the characteristic function of a homogeneous ideal, $\delta(r)$ increases indefinitely with r.

Now take for Θ the dual Picard variety of a Hodge manifold V, and for Z the image $Z=F(V)$ of V by the canonical mapping F of V into Θ; Z is connected, hence it is an infinite set unless it is reduced to a point, in which case we must have $q=0$; this trivial case will be excluded henceforward. The product $\vartheta\vartheta'$ of two theta-functions cannot be 0 on Z unless ϑ or ϑ' is 0 on Z; for, if $\vartheta\vartheta'$ is 0 on Z, $\vartheta[\hat{F}(\hat{M})]\vartheta'[\hat{F}(\hat{M})]$ is 0 on $\hat{V}$, and, as this is the product of two holomorphic functions on $\hat{V}$, one factor must be 0 on an open set, hence everywhere. The integer m and the set (ϑ_μ) having been chosen as above, we use Z, as above, to define a homogeneous polynomial ideal; and we choose r so that $\delta(r)\geqq q+1$.

15. Let E'' be the sum of two Riemann forms for the product $\Theta''=\Theta\times\Theta'$ of Θ and of the dual variety Θ'. In our usual notation, we have $\Theta=A/\Delta$, $\Theta'=A'/\Delta'$, hence $\Theta''=A''/\Delta''$ with $A''=A\times A'$, $\Delta''=\Delta\times\Delta'$; as usual, write A_0, A'_0, and $A''_0=A_0\times A'_0$ for the underlying real spaces of A, A', A''. As E'' is a bilinear form on $A''_0\times A''_0$, it can be written, more explicitly, as $E''(x,x';y,y')$, with $(x,x')\in A_0\times A'_0$, $(y,y')\in A_0\times A'_0$. Then, for each integer $N\geqq 0$, the bilinear form

$$E''_N(x,x';y,y')=E''(x,x';y,y')+NE(x,y)$$

is a Riemann form for Θ''; let Λ_N be the vector-space of all theta-functions belonging to Θ'' and E''_N. Take any $u'\in\Theta'$; we shall now prove that the codimension of the subspace of Λ_{mr}, consisting of the functions in it which are 0 on $Z\times u'$, is $\geqq q+1$.

In fact, as E'' is the sum of two Riemann forms, the result in no. 13 shows that we can choose a function ϑ'_0 in Λ_0 which is not 0 at (z,u'), z being an arbitrarily chosen point in Z. Now, Φ being a homogeneous polynomial of degree r in $U_0,\cdots,U_M$, put

$$\vartheta''_\Phi(x,x')=\vartheta'_0(x,x')\Phi[\vartheta_0(x),\cdots,\vartheta_M(x)];$$

the mapping $\Phi \to \vartheta''_\Phi$ is a linear mapping of the space of such polynomials into Λ_{mr}; and ϑ''_Φ is 0 on $Z \times u'$ if and only if $\Phi(\vartheta_0, \cdots, \vartheta_M)$ is 0 on Z. As the codimension of the inverse image of a linear subspace by a linear mapping is at most equal to that of the subspace, it follows that the subspace of Λ_{mr}, consisting of the functions in it which are 0 on $Z \times u'$, has at least the codimension $\delta(r)$.

16. Now we shall prove that there is a function in Λ_{mr} which is not 0 on any of the sets $Z \times u'$, where u' is any point of Θ'. In fact, let $(\vartheta''_\rho)_{0 \leqq \rho \leqq R}$ be a basis for Λ_{mr}; writing any function ϑ'' in Λ_{mr} as $\vartheta''_t = \sum_\rho t_\rho \vartheta''_\rho$, we can identify Λ_{mr} with the affine space $\boldsymbol{C}^{R+1}$ of dimension $R+1$ over $\boldsymbol{C}$; from this we can derive, in the usual manner, a projective space P^R of dimension R, with the homogeneous coordinates $(t_0, \cdots, t_R)$. The mutual ratios of the ϑ''_ρ are meromorphic functions on Θ'', hence algebraic functions on $\Theta'' = \Theta \times \Theta'$ when Θ, Θ' are identified with algebraic varieties by Lefschetz's theorem (cf. no. 6); therefore the condition that the function ϑ''_t should be 0 at a point (u, u') of Θ'' determines an algebraic set of points $((u, u'), (t_0, \cdots, t_R))$, i. e. a bunch of varieties, on the product-variety $\Theta'' \times P^R$. Hence, for any $z \varepsilon Z$, the condition that ϑ''_t should be 0 at (z, u') determines an algebraic set B_z in the product $\Theta' \times P^R$. Let B be the intersection of all the sets B_z, for all $z \varepsilon Z$; this is again an algebraic set; let B' be its projection on P^R. If $t = (t_0, \cdots, t_R)$ is not in B', then, to every $u' \varepsilon \Theta'$, there is a $z \varepsilon Z$ such that (u', t) is not in B_z, i. e. such that ϑ''_t is not 0 at (z, u'); hence ϑ''_t has the desired property. What we have to show is therefore that $B' \neq P^R$; and for this it is enough to show that no component of B can have a dimension $\geqq R$. Assume therefore that some component W of B has a dimension $\geqq R$; then, if u' is in the projection of W on Θ', the intersection $B \cap (u' \times P^R)$ has a component of dimension $\geqq R - q$. But this intersection is nothing else than the set of points (u', t) of $u' \times P^R$ such that ϑ''_t is 0 on $Z \times u'$; as such functions ϑ''_t form a subspace of Λ_{mr} of codimension $\geqq q + 1$, it follows that $B \cap (u' \times P^R)$ is a linear subspace of $u' \times P^R$ of dimension $\leqq R - q - 1$, and this contradicts the above assumption.

17. After these preliminaries, we are now ready to attack the main problem of this §, which is to construct an algebraic family of divisors on V, parametrized by Θ', containing one and only one representative of each class in $\mathfrak{G}_a/\mathfrak{G}_l$.

In the first place, consider on $A''_0 \times A''_0$ the bilinear form

$$E''_0(x, x'; y, y') = \langle y', x \rangle - \langle x', y \rangle.$$

The dual of $A''_0 = A_0 \times A'_0$ being $A'_0 \times A_0$, this form can be written as $\langle E''_0 x'', y'' \rangle$, where $x'' = (x, x')$, $y'' = (y, y')$, and where E''_0 is the linear mapping $E''_0 = \begin{pmatrix} 0 & -I' \\ I & 0 \end{pmatrix}$ of $A_0 \times A'_0$ onto $A'_0 \times A_0$, I and $I' = {}^tI$ denoting again the identical automorphisms of A_0, A'_0. The scalar multiplication by i in $A'' = A \times A'$ induces on $A''_0 = A_0 \times A'_0$ the automorphism $J'' = \begin{pmatrix} J & 0 \\ 0 & J' \end{pmatrix}$. Then we have $E''_0(J''x'', y'') = \langle y', Jx \rangle + \langle x', Jy \rangle$; as this is symmetric, it follows that, as soon as the integers N, N' are large enough, the bilinear form

$$E''_{N,N'}(x'', y'') = E''_0(x'', y'') + N \cdot E(x, y) + N' \cdot E'(x', y')$$

is a Riemann form for $\Theta'' = \Theta \times \Theta'$, if E, E' are, as before, Riemann forms for Θ, Θ' respectively.

By no. 13, as soon as $N \geqq 2$, there is a theta-function ϑ, belonging to Θ and NE, which is not 0 on $Z = F(V)$. Also, $E(x, y) + E'(x', y')$ is a Riemann form for Θ''; hence, if $E''_{N,N'}$ is a Riemann form for Θ'', $E''_{N+1,N'+1}$ is the sum of two Riemann forms for Θ''; then the main result of no. 16, applied to this instead of E'', shows that there is a theta-function ϑ'', belonging to Θ'' and to $E''_{N+1+mr,N'+1}$, which is not 0 on any of the sets $Z \times u'$. If we write again N, N' instead of $N + 1 + mr$, $N' + 1$, we have thus shown that, for a suitable choice of N, N', there will be a theta-function ϑ, belonging to Θ and NE, which is not 0 on $F(V)$, and a theta-function ϑ'', belonging to Θ'' and $E''_{N,N'}$, which is not 0 on any set $F(V) \times u'$.

In this, it is understood that the periodicity conditions for these functions have been normalized as explained in no. 13; this implies that, if E^0, E'^0 are the symmetric forms attached, as there, to E, E', the symmetric form attached to $E''_{N,N'}$ is:

$$\langle y', x \rangle + \langle x', y \rangle + NE^0(x, y) + N'E'^0(x', y').$$

18. ϑ, ϑ'' being chosen as explained in no. 17, we now put

$$\eta(x, x') = \vartheta(x)\vartheta''(x, x')^{-1}.$$

This is a meromorphic function with the periodicity properties

$$\eta(x + r, x') = \eta(x, x') e(\tfrac{1}{2}\langle x', r \rangle + \tfrac{i}{2}\langle x', Jr \rangle),$$

$$\eta(x, x' + r') = \eta(x, x') e[L_{r'}(x, x')],$$

for $x \varepsilon A$, $x' \varepsilon A'$, $r \varepsilon \Delta$, $r' \varepsilon \Delta'$, $L_{r'}(x, x')$ being a certain complex-linear function on $A \times A'$, which depends upon r'. Hence, if we put, for any $a' \varepsilon A'$,

$$\psi_{a'}(x) = \eta(x, a') e(\tfrac{1}{2}\langle a', x\rangle - \tfrac{1}{2}\langle a', Jx\rangle),$$

we have

$$\psi_{a'}(x + r) = \psi_{a'}(x) e(\langle a', r\rangle),$$

$$\psi_{a'+r'}(x) = \psi_{a'}(x) e[L_{a',r'}(x)],$$

where $L_{a',r'}(x)$ is a complex-linear function on A, depending upon a', r'; $\psi_{a'}(x)$ is a meromorphic function on A, which, by the properties which we have assumed for ϑ and ϑ'', is not everywhere 0 or ∞ on $\hat{F}(\hat{V})$; and it is a multiplicative function, with the multiplicator-set $e(\langle a', r\rangle)$ of absolute value 1. This shows that $e[L_{a',r'}(x)]$ has the multiplicator-set 1, hence it is a constant (as could also readily have been verified by its explicit calculation). Therefore we have

$$\psi_{a'+r'}(x) = c(a', r')\psi_{a'}(x),$$

where $c(a', r')$ depends only upon a' and r'.

19. It follows from the relations in no. 18 that $\psi_{a'}[\hat{F}(\hat{M})]$ is a meromorphic multiplicative function, not everywhere 0 or ∞, on $\hat{V}$, with the multiplicator-set $\mu_{a'} \circ \lambda$, where $\mu_{a'}(r) = e(\langle a', r\rangle)$, and λ is as in no. 7. As $\psi_{a'}$ is multiplied by a constant when a' is replaced by $a' + r'$, its divisor $(\psi_{a'})$ on V depends only upon the image $u' = \omega(a')$ of a' by the canonical homomorphism ω of A' onto $\mathfrak{G}' = A'/\Delta'$; hence we may write it as $(\psi_{a'}) = D(u')$.

$D(u')$ is also the divisor of the multiplicative function

$$\vartheta[\hat{F}(\hat{M})]\vartheta''[\hat{F}(\hat{M}), a']^{-1}$$

on $\hat{V}$. Now $\vartheta[\hat{F}(\hat{M})]$ is a holomorphic function on $\hat{V}$, which, when $\hat{M}$ is replaced by $\sigma\hat{M}$, is multiplied by an exponential factor; hence (cf. no. 10) one can define its divisor W_0 on V. Similarly one can define the divisor W of $\vartheta''[\hat{F}(\hat{M}), x']$ on $V \times \mathfrak{G}'$. By purely local considerations, using the expression of this function as a product of irreducible holomorphic functions in the neighborhood of a point, one shows that the divisor of $\vartheta''[\hat{F}(\hat{M}), a']$ on V, for a fixed a', is the intersection of W with $V \times u'$, where $u' = \omega(a')$, each component of that intersection being counted with its multiplicity; more correctly it is the divisor $W(u')$ on V such that $W(u') \times u'$ is the intersection of W and $V \times u'$. There is, in a case such as this, no difficulty in defining the intersection-multiplicities, by function-theoretic or topological means.

Therefore we have $D(u') = W_0 - W(u')$; and what we have done shows that this is a divisor in the group $\mathfrak{G}_a$, which has the image u' in Θ' by the canonical homomorphism of $\mathfrak{G}_a$ onto Θ'. As Θ' is an algebraic variety, we may express the relation $W(u') \times u' = W \cdot (V \times u')$, where the $\cdot$ denotes the intersection, by saying that the $W(u')$ are an algebraic family of positive divisors, parametrized by the variety Θ'. The canonical homomorphism of $\mathfrak{G}_a$ onto Θ', with the kernel $\mathfrak{G}_l$, can then be defined by saying that it maps each divisor $X \varepsilon \mathfrak{G}_a$ onto the (uniquely determined) point u' of Θ' such that $W(u')$ is linearly equivalent to $W_0 - X$. This implies of course that $W(0)$ is linearly equivalent to W_0, so that, in this last statement, W_0 may be replaced by $W(0)$.

§ III. The main theorem.

20. By an analytic subset of a complex-analytic manifold, we understand a closed subset of it which, in some neighborhood of everyone of its points, is the set of common zeros of a finite number of holomorphic functions; an analytic subset which is irreducible (i. e. which is not the union of two other such sets) is called a subvariety. By a divisor X, we understand, as above, any formal sum of subvarieties of maximal dimension, with integral coefficients; if the latter are positive, X is said to be positive, and we write $X \succ 0$; we write $X \succ Y$ for $X - Y \succ 0$. By the carrier $|X|$ of X, we understand the union of the components of X.

In all this §, we shall denote by V a connected compact complex-analytic manifold of complex dimension n. If S is a complex-analytic manifold, and X a divisor on $V \times S$, the relation $X(s) \times s = X \cdot (V \times s)$ defines a divisor $X(s)$ on V for every $s \varepsilon S$ such that $V \times s$ is not contained in $|X|$; as $|X|$ is closed, the set S' of these points is an open subset of S, which may be considered as a complex-analytic manifold; we shall say that the set of all divisors $X(s)$, for $s \varepsilon S'$, is an *analytic family* of divisors on V, parametrized by S', or having S' as its parameter manifold.

Notations being as above, one may always assume that X has no component of the form $V \times T$, where T is a subvariety of S of maximal dimension; in fact, adding such a component to X does not modify $X(s)$ when s is not on T, and makes $X(s)$ undefined if s is on T. That being assumed, we shall show that the set $S - S'$ of points s of S for which $X(s)$ is not defined is an analytic subset of S, of (complex) dimension $\leqq p - 2$ if p is the dimension of S. It is clearly enough to prove this for $X \succ 0$. Each point $M \times s_0$ of $V \times S$ has a neighborhood where X can be written as

the divisor (Φ) of some holomorphic function Φ; let $z_1, \cdots, z_n$ be local coordinates on V in a neighborhood of M; then, if a point s, sufficiently near to s_0, is such that $V \times s \subset |X|$, we must have $\Phi(z,s) = 0$ identically in z near $z = 0$; therefore the points of $S - S'$ near s_0 are common zeros of all coefficients of Φ when Φ is expanded into a power-series in $z_1, \cdots, z_n$ with coefficients in the ring of holomorphic functions of s at $s = s_0$; conversely, if a point s, sufficiently near to s_0, is a common zero of all these coefficients, then it is easily seen, by analytic continuation along $V \times s$, that $V \times s$ is contained in $|X|$. Applying a theorem of H. Cartan,[5] one sees that $S - S'$ can be defined, in a sufficiently small neighborhood of s_0, by equating to 0 a finite number of the coefficients of Φ; therefore it is an analytic set. If now T is any component of $S - S'$, $V \times T$ is contained in $|X|$; therefore, if T were of dimension $p - 1$, $V \times T$ would be a component of X, which is against our assumption.

In particular, if S is connected, S' also is connected. Then the homological theory of subvarieties of a complex-analytic manifold ([4]) shows that any two divisors of the form $X(s)$ are homologous to each other on V; in particular, if $X(s)$ is homologous to 0 for some s, the same is true for all $s \in S'$.

The main purpose of this § is to show that, *if V is a Hodge manifold, the canonical mapping c of the group $\mathfrak{G}_a$ of the divisors on V, homologous to* 0 *on V, into the Picard variety $\mathfrak{G}'$ of V* (§ I, no. 12) *induces an analytic mapping into $\mathfrak{G}'$ of every analytic family of divisors belonging to $\mathfrak{G}_a$ on V.* In other words, if S, X, S' and $X(s)$ are as above, and if $X(s)$ is homologous to 0 for all $s \in S'$, the mapping $s \rightarrow c[X(s)]$ is a complex-analytic mapping of S' into $\mathfrak{G}'$. Assuming for a moment that this is so, we shall show that *this mapping can be extended to a complex-analytic mapping of S into $\mathfrak{G}'$.* In fact, write as before $\mathfrak{G}' = A'/\Delta'$, where A' is a complex vector-space and Δ' a discrete subgroup of A'; if x' is a variable point of A', we may consider dx' as a vector-valued differential form on A' (with values in A'); as this form is invariant by translations in A' and in particular by Δ', it induces on $\mathfrak{G}'$ a vector-valued form which, in terms of the complex structure of $\mathfrak{G}'$, is a holomorphic form; its inverse image ω on S' by the mapping $s \rightarrow c[X(s)]$ is therefore a holomorphic form on S'. As $S - S'$ is an analytic subset of S and has no component of dimension $p - 1$, it follows, by Hartogs' theorem, that ω can be extended to a holomorphic form on S. As every 1-dimensional cycle on S can be deformed into one on S', the periods of ω on S are the

[5] "Théorème α" of [1], p. 191.

same as those on S'; as the latter must belong to Δ', the same is therefore true on S; as in § I, one concludes from this that the mapping $\hat{s} \rightarrow \int_{\hat{a}}^{\hat{s}} \omega$ of the universal covering of S into A' (for a fixed $\hat{a}$) induces a holomorphic mapping of S into $\mathfrak{G}'$, which, on S', coincides (up to a translation) with the given one $s \rightarrow c[X(s)]$. This proves our assertion.

Even so, our result will remain incomplete in one important respect: we shall not prove it for families of divisors which are parametrized by algebroid varieties. By an algebroid variety, we understand one on which each point has a neighborhood isomorphic to a locally irreducible analytic subset of a complex-analytic manifold. It seems likely that the proof which is to be given here of our main theorem could be applied to this more general problem with no more than slight modifications, or at any rate that the more general result could be shown to follow from ours, if only the foundations of algebroid geometry were not even now as shaky as those of algebraic geometry once used to be.[6]

21. Let $X(s)$ and $Y(t)$ be two analytic families of divisors on V, respectively parametrized by the manifolds S and T. We shall now prove that *the set of points* (s,t) *of* $S \times T$ *such that* $X(s) \succ Y(t)$ *is an analytic subset of* $S \times T$. If we put $\mathfrak{X} = X \times T$, $\mathfrak{X}$ is a divisor on $V \times (S \times T)$, and we have $\mathfrak{X}(s,t) = X(s)$; operating similarly on Y, we see that it is enough to show that, *if* $X(s)$, $Y(s)$ *are two families, parametrized by the same manifold* S, *the set of points* $s \in S$ *such that* $X(s) \succ Y(s)$ *is an analytic subset of* S. If we write $X = X' - X''$, $Y = Y' - Y''$, where X', X'', Y', Y'' are positive divisors, $X(s) \succ Y(s)$ is equivalent to $X'(s) + Y''(s) \succ X''(s) + Y'(s)$; therefore it will be enough to prove our assertion for positive X and Y. As the question is purely local so far as S is concerned, we may assume that S is a neighborhood of 0 in the space of p complex variables.

Let M be any point on V; take local coordinates $(w, z_1, \cdots, z_{n-1})$ on V in a neighborhood of M, such that the subvariety (of complex dimension 1) of that neighborhood defined by $z_1 = \cdots = z_{n-1} = 0$ is not contained in $|X(0)|$ nor in $|Y(0)|$. Then, by Weierstrass's lemma, X can be written in a neighborhood of $M \times 0$ on $V \times S$ as the divisor (Φ) of a function Φ of the form

[6] It is to be hoped that someone will soon undertake the taxing but necessary task of consolidating them or rather building them up anew. As some at least of the main difficulties have been removed by the recent work of H. Cartan, W. L. Chow, N. Hamilton, K. Oka and others, the time seems ripe for such an undertaking.

$$\Phi(w, z, s) = w^a + \sum_{i=1}^{a} w^{a-i}\phi_i(z, s),$$

where the ϕ_i are holomorphic and 0 at the origin; similarly, Y can be written as (Ψ) in a neighborhood of $M \times 0$, with Ψ of the form

$$\Psi(w, z, s) = w^b + \sum_{j=1}^{b} w^{b-j}\psi_j(z, s),$$

where the ψ_j are holomorphic and 0 at the origin. Considering Φ and Ψ as polynomials over the ring of holomorphic functions of z, s at the origin, we can divide Φ by Ψ; let R be the remainder, which is of the form $R(w, z, s) = \sum_{h=0}^{b-1} w^h \chi_h(z, s)$, where the χ_h are holomorphic at the origin. Put $|z| = \sup_{1 \leq i \leq n-1} |z_i|$; take $\rho > 0$ and an open neighborhood S' of 0 on S so that the range of the local coordinates (w, z) contains the closure of the polycylinder P defined by $|w| < \rho$, $|z| < \rho$, that Φ, Ψ are holomorphic in $P \times S'$, and that X, Y coincide respectively with (Φ) and with (Ψ) in $P \times S'$. Take $\epsilon > 0$ and $< \rho$ and a neighborhood S'' of 0 in S' so that, for $|z| < \epsilon$ and $s \in S''$, Φ and Ψ, considered as polynomials in w, have all their roots inside the disk $|w| < \rho$. Then, if $\zeta = (\zeta_1, \cdots, \zeta_{n-1})$ is such that $|\zeta| < \epsilon$, and if $s \in S''$, the roots of $\Phi(w, \zeta, s) = 0$, counted with their multiplicities, make up the intersection of $X(s)$ with the subvariety of P (of complex dimension 1) defined by $z = \zeta$; and the roots of $\Psi(w, \zeta, s) = 0$ make up the intersection of Y with that subvariety; therefore, if $X(s) \succ Y(s)$ for some $s \in S''$, $\Phi(w, \zeta, s)$ must be a multiple of $\Psi(w, \zeta, s)$, i. e. $R(w, \zeta, s)$ must be 0, whenever $|\zeta| < \epsilon$; for such an s, each coefficient $\chi_h(z, s)$ of R must therefore be identically 0 as a function of z; if the $\chi_h(z, s)$ are expanded into power-series in $z_1, \cdots, z_{n-1}$ with coefficients which are holomorphic functions of s in S', every $s \in S''$ such that $X(s) \succ Y(s)$ must be a common zero of all these coefficients. Conversely, if s is such a common zero, then $R = 0$, and so Φ is a multiple of Ψ in the ring of polynomials over holomorphic functions of z in $|z| < \rho$ and a fortiori in the ring of holomorphic functions in P. Therefore, when that is so, no component of $X(s) - Y(s)$ which has a coefficient < 0 in that divisor can have any point in P.

To every point M on V, we have thus assigned a neighborhood $P = P(M)$ of M with the above stated properties; V can then be covered with a finite number of such neighborhoods $P_\nu = P(M_\nu)$. To each M_ν, the above construction also assigns a neighborhood S''_ν of 0 on S; we may replace the S''_ν by their intersection S''. Then our construction defines infinitely many

functions $f_\alpha(s)$, holomorphic in S'', viz. the coefficients of the expansions into power-series in z of the functions $\chi_h(z,s)$ attached to the various points M_ν, with the following properties: if $s \in S''$ is such that $X(s) \succ Y(s)$, then all the $f_\alpha(s)$ are 0; conversely, if $s \in S''$ and all the $f_\alpha(s)$ are 0, then no component of $X(s) - Y(s)$ with a coefficient < 0 in that divisor can have a point in any P_ν; as the P_ν are a covering of V, this implies that $X(s) \succ Y(s)$. By Cartan's theorem,[5] this completes our proof.

If $X(s)$, $Y(t)$ are again two analytic families of divisors on V, then *the set of points* (s,t) *of* $S \times T$ *such that* $X(s) = Y(t)$ *is an analytic subset of* $S \times T$; in fact, it is the intersection of the two analytic subsets determined by $X(s) \succ Y(t)$ and by $Y(t) \succ X(s)$.

22. If X is any divisor on V, we shall denote by $L(X)$ the set of all meromorphic functions ϕ on V which are either identically 0 or such that $(\phi) \succ -X$; this is a vector-space over the field of constants. We shall now prove that, *for every divisor* X *on* V, *the vector-space* $L(X)$ *is of finite dimension.*

If $X = X' - X''$, with positive X' and X'', we have $L(X) \subset L(X')$; so it will be enough to prove our theorem for a positive divisor X. Then every point M on V has a neighborhood $U = U(M)$ in which X can be written as the divisor (Φ_M) of a function Φ_M, holomorphic in U; and the set $\Phi_M \cdot L(X)$ of all functions $\psi = \Phi_M\phi$, with $\phi \in L(X)$, consists of functions which are holomorphic in U. Consider the ideal generated by $\Phi_M L(X)$ in the ring of holomorphic functions at M; by Cartan's theorem,[5] this ideal is generated by a finite number of elements $\psi_i^{(M)}$ of $\Phi_M L(X)$, and there is an open neighborhood $U' = U'(M)$ of M and a constant $k = k(M)$ with the following properties: the closure $\bar{U}'$ of U' is contained in U; and every $\psi \in \Phi_M L(X)$ such that $|\psi| \leqq 1$ in $\bar{U}$ can be expressed as $\psi = \sum_i \lambda_i \psi_i^{(M)}$, where the λ_i are holomorphic and such that $|\lambda_i| \leqq k$ on $\bar{U}'$. Take a covering of V by a finite number of neighborhoods $U'_\nu = U'(M_\nu)$; to each there belongs a function $\Phi_\nu = \Phi_{M_\nu}$, a constant $k_\nu = k(M_\nu)$, and a finite set of functions $\psi_i^{(M_\nu)} \in \Phi_\nu L(X)$; put $k = \sup_\nu k_\nu$, and call $\phi_1, \cdots, \phi_m$ the functions $\Phi_\nu^{-1}\psi_i^{(M_\nu)}$; these are functions in $L(X)$. Put $C_\nu = \bar{U}'_\nu$. Then, for every $\phi \in L(X)$ and every ν, there are functions $\lambda_{\nu j}$ holomorphic on C_ν, such that $\phi = \sum_{j=1}^{m} \lambda_{\nu j}\phi_j$ on C_ν; furthermore, if $|\Phi_\nu\phi| \leqq 1$, on C_ν, the $\lambda_{\nu j}$ may be taken such that $|\lambda_{\nu j}| \leqq k$ on C_ν.

Now put

$$\| \phi \| = \sup_\nu \sup_{C_\nu} | \Phi_\nu \phi |$$

for each $\phi \in L(X)$; then every $\phi \in L(X)$ can be written in C_ν as $\phi = \sum_j \lambda_{\nu j}\phi_j$, with $|\lambda_{\nu j}| \leqq k \, \| \phi \|$ on C_ν. We shall now prove that, if the vector-space $L(X)$ is considered as a normed vector-space with the norm $\| \phi \|$, it is locally compact, i. e. that the "unit-sphere" $\| \phi \| \leqq 1$ is compact. In fact, take any sequence ϕ_α of functions in $L(X)$ such that $\| \phi_\alpha \| \leqq 1$; in C_ν, ϕ_α can be written as $\phi_\alpha = \sum_j \lambda_{\nu j}{}^{(\alpha)}\phi_j$, with $|\lambda_{\nu j}{}^{(\alpha)}| \leqq k$ on C_ν. By the well-known elementary theorem on sequences of bounded holomorphic functions, we may, after replacing the sequence ϕ_α by a suitable subsequence, assume that, for each ν and each j, the sequence $\lambda_{\nu j}{}^{(\alpha)}$ converges to a limit $\lambda_{\nu j}$ uniformly on every closed subset of U'_ν. Then one sees immediately that there is a function $\phi \in L(X)$ such that $\phi = \sum \lambda_{\nu j}\phi_j$ in U'_ν for each ν; furthermore, as we have

$$\Phi_\nu\phi = \sum_j \lambda_{\nu j}(\Phi_\nu\phi_j) = \lim_{\alpha\to\infty} \sum_j \lambda_{\nu j}{}^{(\alpha)}\Phi_\nu\phi_j = \lim_{\alpha\to+\infty} (\Phi_\nu\phi_\alpha)$$

uniformly on every closed subset of U'_ν, we see that the sequence ϕ_α tends to ϕ in the topology defined by the norm $\| \phi \|$, and that $\| \phi \| \leqq 1$. This shows that the normed vector-space $L(X)$ is locally compact; by a well-known elementary theorem of Banach, this implies that it has a finite dimension.

23. From now on, we assume that V is a Hodge manifold; and we proceed to prove the theorem stated in no. 20. Let $X(s)$ be an analytic family of divisors, parametrized by a manifold S; we have to prove that, if the divisors $X(s)$ are homologous to 0, $s \to c[X(s)]$ is a complex-analytic mapping of S into Θ'. We may assume that S is connected; and we choose a point s_0 on S. Call X', $-X''$ the positive and negative parts of X, i. e. the sums of terms, with positive and negative coefficients respectively, in the expression of X in terms of its irreducible components. Then $X(s)$ is the sum of the divisors $X'(s) - X'(s_0)$, $-[X''(s) - X''(s_0)]$ and $X(s_0)$, which are all homologous to 0, and it is enough to prove our assertion for the families $X'(s) - X'(s_0)$, $X''(s) - X''(s_0)$. In other words, it will be enough to show that if $X(s)$ is any family of positive divisors parametrized by the connected manifold S, $s \to c[X(s) - X(s_0)]$ is an analytic mapping of S into Θ'. Now, in § II, we have constructed an analytic family $W(u')$ of positive divisors on V, parametrized by Θ', such that $u' = c[W(0) - W(u')]$. This implies that there is one and only one point u' on Θ', given by $u' = c[X(s) - X(s_0)]$, for which $W(0) - W(u')$ is linearly equivalent to $X(s) - X(s_0)$, i. e. $W(u') + X(s)$ to $W(0) + X(s_0)$. Let $\phi_0, \cdots, \phi_m$ be a basis for the vector-

space $L[W(0)+X(s_0)]$; then, if s and u' are as we have said, there is one and only one point ξ in the projective m-dimensional space P^m such that, if $(\xi_0,\cdots,\xi_m)$ is a set of homogeneous coordinates for ξ, the divisor of the function $\phi=\sum_{i=0}^{m}\xi_i\phi_i$ is

$$(\phi)=W(u')+X(s)-[W(0)+X(s_0)].$$

For every $\phi=\sum_i\xi_i\phi_i$ in $L[W(0)+X(s_0)]$, put $Z(\xi)=(\phi)+W(0)+X(s_0)$, where ξ denotes the point with the homogeneous coordinates $(\xi_0,\cdots,\xi_m)$ in P^m; one sees at once that $Z(\xi)$ is an analytic family of divisors parametrized by P^m. By the final result of no. 21, the set of all points (s,u',ξ) in $S\times\Theta'\times P^m$ such that $Z(\xi)=W(u')+X(s)$ is an analytic subset Γ of $S\times\Theta'\times P^m$. But we have just shown that, for every $s\in S$, there is one and only one point (u',ξ) in $\Theta'\times P^m$ such that (s,u',ξ) is in Γ; as Γ is closed and Θ' and P^m are compact, this implies in the first place that Γ is the graph of a continuous mapping f of S into $\Theta'\times P^m$; it is then easily seen that f induces a holomorphic mapping on every subvariety of complex dimension 1 of any open subset of S; from this, using elementary results on functions of several complex variables, one deduces easily that f is holomorphic on S itself. This completes the proof.

The same method can be applied to the determination of all positive divisors in a given homology class. Consider a homology class which contains at least one positive divisor X_0; let $\phi_0,\cdots,\phi_m$ be a basis for the vector-space $L[W(0)+X_0]$; for every $\phi=\sum_i\xi_i\phi_i$ in that space, put $Z(\xi)=(\phi)+W(0)+X_0$, ξ being the point in P^m with the homogeneous coordinates $(\xi_0,\cdots,\xi_m)$. Then we see, as above, that for every positive divisor X homologous to X_0 there is one and only one point (u',ξ) of $\Theta'\times P^m$ such that $Z(\xi)=W(u')+X$; therefore this formula determines a one-to-one correspondence between such divisors X and the points (u',ξ) of $\Theta'\times P^m$ such that $Z(\xi)\succ W(u')$. As we have shown, the latter relation determines an analytic subset of $\Theta'\times P^m$, which, since Θ' and P^m are compact, is the union of a finite number of subvarieties of $\Theta'\times P^m$; moreover, as Θ' is an algebraic variety, these are algebraic subvarieties of $\Theta'\times P^m$ by Chow's theorem ([2]). This shows that *the positive divisors in a given homology class on V make up a finite number of algebraic families*, i. e. of families parametrized by algebraic varieties; of course the latter may have singular points and so need not be complex-analytic manifolds, but are algebroid varieties in the sense defined at the end of no. 20. One may also observe that for a given u' in Θ' the points ξ of P^m

such that $Z(\xi) \succ W(u')$ are those in a linear subvariety of P^m, corresponding to the subspace $L[W(0) + X_0 - W(u')]$ of the vector-space $L[W(0) + X_0]$; in particular, each one of the above families of positive divisors may be considered as a fibre-variety over a subvariety of $\mathfrak{G}'$, the fibres being projective spaces.

§ IV. The higher Jacobian varieties.

24. I shall take this opportunity for making explicit some results implicitly contained in Hodge's book ([5]; cf. also [6]), as this can be done very simply in the language which has been introduced above in § I and throws some additional light on the results of that §.

As before, let V be a Kähler manifold of complex dimension n; take p such that $0 \leqq p \leqq n-1$; call now A_0 and A'_0 the homology groups of V over the real number-field $\boldsymbol{R}$ for the dimensions $2p+1$ and $2n-2p-1$ respectively. These are vector-spaces of finite dimension over $\boldsymbol{R}$; call Δ and Δ' the subgroups of A_0 and of A'_0 consisting of those homology classes which contain cycles with integral coefficients (i. e. the images in A_0 and in A'_0 of the homology groups of V with integral coefficients for the same dimensions); Δ is a discrete subgroup of A_0 of maximal rank, i. e. of rank equal to the dimension of A_0, so that A_0/Δ is a torus of that same dimension; and the same is true of A'_0 and Δ'.

By the Poincaré duality theorem, A_0 and A'_0 are put into duality with each other by the intersection-number (or "Kronecker index"), which induces a bilinear form $\langle x', x\rangle$ on $A'_0 \times A_0$; moreover, by the same theorem, Δ and Δ' are associated to each other in that duality, i. e. Δ' consists of all the elements r' of A'_0 such that $\langle r', r\rangle \equiv 0 \bmod 1$ for all $r \varepsilon \Delta$. We may therefore identify A'_0 with the dual space to A_0, i. e. with the cohomology group of V of dimension $2p+1$ with real coefficients, and similarly A_0 may be identified with the cohomology group of V of dimension $2n-2p-1$ with real coefficients. In particular, A_0 and A'_0 may be identified with the de Rham groups of degrees $2n-2p-1$ and $2p+1$ respectively; by the de Rham group of degree d, we understand the group of closed differential forms of degree d on V modulo the exterior differentials of forms of degree $d-1$. Then Δ, Δ' are the classes of forms of degrees $2n-2p-1$ and $2p+1$ with integral periods; and if elements x, x' of A_0, A'_0 are the cohomology classes of two closed forms ω, ω' of respective degrees $2n-2p-1$ and $2p+1$, we have $\langle x', x\rangle = \int_V \omega'\omega$.

By Hodge's existence theorems, each cohomology class contains one and only one harmonic form; so A_0 may also be identified with the space of harmonic forms of degree $2n-2p-1$. As in no. 7, consider the operator C derived from the complex structure of V; if ω is a form of degree d, we have $C(C\omega)=(-1)^d\omega$, so that C induces an operator such that $C^2=-I$ on forms of any odd degree. As C commutes with the operator $\Delta=d\delta+\delta d$, it transforms harmonic forms into harmonic forms; call J the operator induced by $-C$ on the space A_0 of harmonic forms of degree $2n-2p-1$; this is an endomorphism of A_0 satisfying $J^2=-I$; therefore A_0 must be of even dimension, and J can be used to define a complex structure on A_0; with that structure, A_0 becomes a vector-space over the field of complex numbers, which will be denoted by A. In a similar manner, A'_0 becomes a vector-space A' over complex numbers by means of the operator J' induced by $-C$ on harmonic forms of degree $2p+1$. Let ω, ω' be two harmonic forms of respective degrees $2n-2p-1$, $2p+1$; as C is an automorphism of the ring of differential forms, we have $C(\omega'\omega)=C\omega'\cdot C\omega$; as every form ζ of degree $2n$ can be written as $fdz_1d\bar{z}_1\cdots dz_nd\bar{z}_n$ in terms of local coordinates, we have $C\zeta=\zeta$ for such a form, and in particular $C(\omega'\omega)=\omega'\omega$. This gives $J'\omega'\cdot J\omega=\omega'\omega$, whence $\langle J'\omega', J\omega\rangle=\langle\omega',\omega\rangle$, or, if we replace ω by $J^{-1}\omega$, $\langle J'\omega',\omega\rangle=\langle\omega', J^{-1}\omega\rangle$; this shows that J' is the transpose of J^{-1} for the duality determined by the bilinear form $\langle x', x\rangle$. According to our definitions in § I, no. 1, this makes A' the "antidual" space of A; and, if we put $\Theta=A/\Delta$, $\Theta'=A'/\Delta'$, these are two complex toruses, dual to each other.

25. We shall now give a definition of J, as an endomorphism of the de Rham group of degree $2n-2p-1$, which depends only upon the complex structure of V and not upon the choice of a Kähler metric; the existence of a Kähler metric on V, however, is essential to our purposes. Take any d such that $0\leqq d\leqq 2n$; in each cohomology class of degree d, there are forms ω such that $d\omega=d(C\omega)=0$; in fact, the harmonic form of that class has that property. We now prove that, for $d\omega=0$ and $d(C\omega)=0$, the homology class of $C\omega$ depends only upon that of ω. In fact, assume that ω is homologous to 0; let ζ be any harmonic form of degree $2n-d$, and put $\zeta'=C^{-1}\zeta$; then ζ' is harmonic; we have $\zeta'\omega=C(\zeta'\omega)$ since $\zeta'\omega$ is of degree $2n$, and $C(\zeta'\omega)=C\zeta'\cdot C\omega=\zeta(C\omega)$, and so $\zeta'\omega=\zeta(C\omega)$. As ω is homologous to 0, we have $\int_V\zeta'\omega=0$; this gives $\int_V\zeta(C\omega)=0$; as this is true for all harmonic forms ζ, i. e. for representatives of all cohomology classes of degree

$2n-d$, and as $C\omega$ is closed, it implies that $C\omega$ is homologous to 0.[7] If now x is a cohomology class of degree d, and ω is a representative of that class such that $d\omega=0$ and $d(C\omega)=0$, we shall denote by Cx the class of the form $C\omega$; this is an endomorphism of the de Rham group of dimension d which satisfies $C^2=(-1)^d I$ and depends only upon the complex structure of V; for $d=2n-2p-1$, the operator J defined above is $J=-C$.

The complex torus Θ defined above by means of the cohomology group A_0 of degree $2n-2p-1$, of the complex structure determined by J, and of the subgroup Δ of classes of A_0 with integral periods, will henceforth be denoted by Θ_p; with this notation, the torus denoted above by Θ' now appears as Θ_{n-p-1}. Instead of A, Δ, we now write $A^{(p)}$, $\Delta^{(p)}$, so that $\Theta_p=A^{(p)}/\Delta^{(p)}$.

Thus, if V is any connected compact complex-analytic manifold of complex dimension n on which there exists at least one Kähler metric, we have defined a sequence $\Theta_0, \Theta_1, \cdots, \Theta_{n-1}$ of complex toruses invariantly attached to it, in such a way that Θ_p and Θ_{n-p-1} are dual to each other for $0 \leqq p \leqq n-1$. The torus Θ attached to V in § I, no. 7, is no other than Θ_0 in this notation, and its dual Θ' is Θ_{n-1}.

26. It will now be shown that, if V is a Hodge manifold, not only Θ_0 and Θ_{n-1} (as we have seen in § I) but all the Θ_p are abelian varieties; these will be called *the Jacobian varieties of* V. As before, denote by Ω the fundamental form of degree 2 of the Kähler metric, which by assumption has integral periods; we have $C\Omega=\Omega$. If x is the cohomology class of a form ω, that of $\Omega\omega$ will be denoted by Lx. If ω is a representative of the class x such that $d\omega=0$ and $d(C\omega)=0$, we have $d(\Omega\omega)=0$ and $d[C(\Omega\omega)]=d[\Omega(C\omega)]=0$, and therefore, according to our definitions, $L(Cx)=C(Lx)$. For forms of odd degree, this gives $L(Jx)=J(Lx)$; so L is a complex-linear mapping of $A^{(p)}$ into $A^{(p-1)}$ for all p. As $\Omega\omega$ has integral periods whenever ω has integral periods, L maps $\Delta^{(p)}$ into $\Delta^{(p-1)}$; so L induces a complex-analytic homomorphism of Θ_p into Θ_{p-1}, which will again be denoted by L. But by

[7] Using de Rham's operator H (cf. [3], § 19), this can be proved more briefly: if $d\omega=0$, $H\omega=0$ is a necessary and sufficient condition for ω to be homologous to 0; as H commutes with C, $H\omega=0$ implies $H(C\omega)=0$. This argument applies also to the projection operators $P_{a,b}$ which map each form ω of degree $a+b$ into the sum of the terms of type (a,b) in the expression of ω in terms of local complex coordinates. As $d\omega=d(C\omega)=0$ implies $d(P_{a,b}\omega)=0$, and as H commutes with $P_{a,b}$, one can define a projection operator $P_{a,b}$, invariantly related to the complex structure on V, on the cohomology group of V with real coefficients for the dimension $a+b$. The dimension of the image of this group under $P_{a,b}$ is Hodge's invariant $\rho^{a,b}$ ([6], Theorem I, p. 109).

Hodge's theorems, if ω is harmonic, so is $\Omega\omega$, and the mapping $\omega \rightarrow \Omega\omega$ of the space of harmonic forms of degree d into the space of harmonic forms of degree $d+2$ has the kernel 0 if $d \leqq n-1$ and maps the former space onto the latter if $d \geqq n-1$. So L maps Θ_p onto Θ_{p-1} for $2p \leqq n$, and it maps Θ_p into Θ_{p-1} with a finite kernel if $2p \geqq n$.

For $d \leqq n-1$, call B, B_1, B_2 the spaces of harmonic forms of degrees d, $d+2$, $2n-d$ respectively. As we have just said, L maps B isomorphically onto its image $L(B)$ in B_1. On the other hand, call E the kernel of the mapping L^{n-1-d} of B_1 into B_2, i. e. of the mapping $\omega \rightarrow \Omega^{n-1-d}\omega$; E is the space of "effective" harmonic forms of degree $d+2$ in Hodge's terminology, and Hodge has proved that B_1 is the direct sum of $L(B)$ and of E. Let Δ, Δ_1, Δ_2 be the groups of forms with integral periods in B, B_1, B_2 respectively; as we have said, L maps Δ into Δ_1, and so L^{n-1-d} maps Δ_1 into Δ_2; taking sets of generators of Δ, Δ_1, Δ_2 as bases for B, B_1, B_2 respectively, we see that the mappings L and L^{n-1-d} will then be given by matrices with integral coefficients, and so the linear subspaces $L(B)$ and E of B_1 can both be defined by linear equations with integral coefficients. This implies that the group $\Delta'_1 = L(\Delta) + (E \cap \Delta_1)$ is a subgroup of B_1 of maximal rank and a subgroup of Δ_1 of finite index, and that $E \cap \Delta_1$ is a subgroup of E of maximal rank.

Now, assuming that d is odd, we give to B, B_1, B_2 the complex structure determined by $J = -C$; as L and L^{n-1-d} are complex-linear mappings, $L(B)$ and E are then complex-linear subspaces of B_1. If we put $d = 2p-1$, we have $2p \leqq n$, and, with the notation explained above, $B = A^{(n-p)}$, $B_1 = A^{(n-p-1)}$, $\Delta = \Delta^{(n-p)}$, $\Delta_1 = \Delta^{(n-p-1)}$, and therefore $\Theta_{n-p} = B/\Delta$, $\Theta_{n-p-1} = B_1/\Delta_1$. Put $\Theta'_{n-p-1} = B_1/\Delta'_1$; the identical mapping of B_1 onto itself maps Δ'_1 into Δ_1, and so it induces a complex homomorphism of Θ'_{n-p-1} onto Θ_{n-p-1}; therefore these two toruses are isogenous (§ I, no. 2); but the former is the direct product of the complex torus $L(B)/L(\Delta)$, which is isomorphic to $B/\Delta = \Theta_{n-p}$, and of the complex torus $\Theta''_{n-p-1} = E/(E \cap \Delta_1)$. So Θ_{n-p-1} is isogenous to the product of Θ_{n-p} and of Θ''_{n-p-1} whenever $2p \leqq n$; and if we prove that Θ''_{n-p-1} is an abelian variety for $2p \leqq n-1$, it will follow by induction on p that the same is true of Θ_{n-p-1} for the same values of p, and also of Θ_p for these values of p since Θ_p is dual to Θ_{n-p-1}; as every integer q such that $0 \leqq q \leqq n-1$ can be written either as $q = p$ or as $q = n-p-1$ with $2p \leqq n-1$, our proof will then be complete. Now consider the bilinear form

$$F(\omega, \omega') = (-1)^{p+1} \int_V \omega\omega'\Omega^{n-2p-1}$$

for $\omega \in E$, $\omega' \in E$; as ω, ω' are of odd degree, we have $\omega\omega' = -\omega'\omega$, and F is alternating. If ω, ω' are in $E \cap \Delta_1$, i. e. if they have integral periods, $F(\omega, \omega')$ is an integer. Finally, again by Hodge's theorems, we have, for every $\omega \in E$:

$$\Omega^{n-2p-1}(C\omega) = (-1)^{p+1}(n-2p-1)\,!\,{*\omega},$$

and therefore

$$F(J\omega, \omega') = (n-2p-1)\,!\int_V \omega'({*\omega})\,;$$

this is symmetric and positive-definite; so F is a Riemann form for $\mathfrak{G}''_{n-p-1}$, and this torus is an abelian variety.

27. By an analytic cycle of (complex) dimension p on V, we understand a finite sum, with integral coefficients, of subvarieties of V of complex dimension p. Let $\mathfrak{Z}_h^{(p)}$ be the group of those cycles of dimension p on V which are homologous to 0 on V (with integral coefficients). Under the assumption that each analytic cycle on V defines a "current," i. e. essentially that the integral of a differential form of degree $2p$ on the manifold of simple points of a subvariety of V of complex dimension p is always convergent, one can define a canonical mapping of the group $\mathfrak{Z}_h^{(p)}$ into the complex torus $\mathfrak{G}_p$. This assumption is "undoubtedly" true, but no proof of it seems to have been published; in order to simplify the definition of the canonical mapping of $\mathfrak{Z}_h^{(p)}$ into $\mathfrak{G}_p$, I shall assume even more, viz. that every cycle in $\mathfrak{Z}_h^{(p)}$ can be expressed as the boundary of a singular chain of such a nature that differential forms of degree $2p+1$ can be integrated on it.

That being assumed, let X be any cycle in $\mathfrak{Z}_h^{(p)}$; put $X = bQ$, where b is the boundary operator and Q is a chain with integral coefficients. Let x' be a cohomology class of degree $2p+1$ with real coefficients; take a representative ω for x' such that $d\omega = 0$ and $d(C\omega) = 0$; we shall prove that $\int_Q \omega$ depends only upon x'. In fact, assume that ω is homologous to 0. As ω is then orthogonal to all harmonic forms, there is a form ζ such that $\omega = \Delta\zeta = d\delta\zeta + \delta d\zeta$; this implies $d(\delta d\zeta) = 0$; as $\delta d(d\zeta) = 0$, $d\zeta$ is therefore harmonic; as it is homologous to 0, it is 0. This shows that $d\zeta = 0$ and $\omega = d\delta\zeta$. As Δ commutes with C, we have $C\omega = \Delta(C\zeta)$; as $d(C\omega) = 0$, we can reason in the same manner on $C\omega$, $C\zeta$ as we have done on ω, ζ, and so we get $d(C\zeta) = 0$ and $C\omega = d\delta C\zeta$ (which proves again that $C\omega$ is homologous to 0). By another fundamental identity of the Kähler metric, we have $\Lambda d - d\Lambda = -C\delta C^{-1}$ (the nature of the operator Λ need not concern us here) and so $dC\delta = -d\Lambda dC$ and $dC\delta\zeta = -d\Lambda(dC\zeta) = 0$. By Stokes's theorem,

we have $\int_Q \omega = \int_X \delta\zeta$. But if λ is any form of degree $2p$, λ and $C\lambda$ induce the same form on every analytic subvariety of V of complex dimension p; in particular, we have $\int_X \delta\zeta = \int_X C\delta\zeta = \int_Q d(C\delta\zeta) = 0$. This shows that $\int_Q \omega$ depends only upon x'; as it is a linear function of x', it can be written as $\langle x', x\rangle$, where $x = x(Q)$ is an element of the homology group A_0 of V with real coefficients for the dimension $2p+1$. If X is given, Q is well determined modulo cycles of dimension $2p+1$ with integral coefficients; if Z is such a cycle, its class z in A_0 belongs to the subgroup of A_0 which we have called Δ, or more precisely $\Delta^{(p)}$; if we replace Q by $Q+Z$, $x(Q)$ is replaced by $x(Q+Z) = x(Q) + z$. Therefore, when X is given, $x(Q)$ is well determined modulo Δ; if we denote by $c_p(X)$ the image of $x(Q)$ in the factor-group A_0/Δ, $X \rightarrow c_p(X)$ is therefore a homomorphism of $\mathfrak{G}_h{}^{(p)}$ into Θ_p whose definition depends only upon the complex structure of V provided there exists at least one Kähler metric on V. This is the canonical mapping that we wished to construct.

For $p=0$, the group $\mathfrak{G}_h{}^{(0)}$ is the group generated by the cycles $M-P$ if P is a fixed point on V and M is arbitrary; in order to determine c_0, it is therefore enough to know the mapping $M \rightarrow c_0(M-P)$ of V into Θ_0; it is easy to see that this is no other than the canonical mapping F which was defined in § I, no. 7. On the other hand, it follows from a theorem of Kodaira (Theorem 3, p. 110, in [3]) that the mapping c_{n-1} of $\mathfrak{G}_h{}^{(n-1)} = \mathfrak{G}_a$ into Θ_{n-1} is the same as the canonical mapping c defined above in § I, no. 12.

It is of course tempting to conjecture that any analytic family of cycles of dimension p on V, belonging to the group $\mathfrak{G}_h{}^{(p)}$, is mapped analytically into Θ_p by c_p; this seems very likely, but I have not yet succeeded in proving it. One would also like c_p to map $\mathfrak{G}_h{}^{(p)}$ onto Θ_p, as this would pave the way to an algebraic interpretation of the Jacobian varieties Θ_p when V is an algebraic variety; in fact, Θ_p could then be described as the factor-group of $\mathfrak{G}_h{}^{(p)}$ modulo a concept of equivalence which one might hope to define in algebraic terms. But Prof. Hodge has pointed out to me that c_p cannot map $\mathfrak{G}_h{}^{(p)}$ onto Θ_p unless all harmonic forms of degree $2p+1$ are sums of forms of type $(p+1, p)$ and of forms of type $(p, p+1)$; as in general this is not the case, some altogether new idea is required if the higher Jacobian varieties are to acquire the place which seems to belong to them in algebraic geometry.

UNIVERSITY OF CHICAGO.

REFERENCES.

[1] H. Cartan, "Idéaux de fonctions analytiques de n variables complexes," *Ann. Sc. E. N. S.* (III) 61 (1944), p. 149.

[2] W. L. Chow, "On compact complex analytic varieties," *American Journal of Mathematics*, vol 71 (1949), p. 893.

[3] G. de Rham and K. Kodaira, *Harmonic Integrals* (Mimeographed notes, The Institute for Advanced Study, Princeton, 1950).

[4] N. Hamilton, *Homology theory of complex-analytic manifolds*, Ph. D. thesis, Chicago.

[5] W. V. D. Hodge, *The theory and applications of harmonic integrals*, Cambridge University Press, Cambridge, 1941.

[6] ———, "A special type of Kähler manifold," *Proc. L. M. S.* (III), 1 (1951), p. 104.

[7] J. Igusa, "On the Picard varieties attached to algebraic varieties," *American Journal of Mathematics*, vol. 74 (1952), pp. 1-22.

[8] S. Lefschetz, "On certain numerical invariants of algebraic varieties with application to Abelian varieties," *Transactions of the American Mathematical Society*, vol. 22 (1921), p. 327.

[9] C. L. Siegel, *Analytic functions of several complex variables* (Mimeographed notes, The Institute for Advanced Study, Princeton, 1949).

[10] A. Weil, "Sur la théorie des formes différentielles attachées à une variété analytique complexe," *Comm. Math. Helv.*, vol. 20 (1947), p. 110.

[11] ———, *Théorèmes fondamentaux de la théorie des fonctions thêta* (mimeographed), Séminaire Bourbaki, Paris, 1949.

ON PICARD VARIETIES.*

By Wei-Liang Chow.

1. Introduction. In a recent paper[1] J. Igusa has developed a theory of the Picard variety attached to a non-singular algebraic variety V. Let $\mathfrak{G}_0(V)$ be the group of all divisors in V which are continuously equivalent to zero, and let $\mathfrak{G}_l(V)$ be the group of all divisors in V which are linearly equivalent to zero; let A be the Abelian variety determined by the periods of the Abelion integrals of the first kind on V, and let P be the dual Abelian variety of A in the sense of Weil (see section 3). The variety P is called the Picard variety attached to V, and Igusa's main result is that the group $\mathfrak{G}_0(V)$ can be mapped homomorphically onto the Picard variety P such that the kernel of this homomorphism is precisely the group $\mathfrak{G}_l(V)$. A. Weil[2] has taken up the same problem (for the possibly more general case of a Hodge variety) by a quite different method and has obtained among others also this result of Igusa, which he calls the first duality theorem of Igusa. There is however one important point concerning this theorem which is not settled in Igusa's paper; this is the question whether the mapping of $\mathfrak{G}_0(V)$ onto P is an analytic mapping in the sense (to be defined more precisely later in section 2) that it is analytic for every analytic system of divisors in $\mathfrak{G}_0(V)$. The importance of this question has been recognized by Weil, who considers the solution of this question, which he has obtained by a very elegant method, as the Main Theorem of his theory.[3] In this paper we shall give a new, simple proof of Igusa's duality theorem, which will at the same time take care of this additional question of analyticity almost without any further effort. Our method, like that of Igusa, is based on the existence of a "sufficiently general" curve in V, but we shall make systematic use of the concept of

* Received February 4, 1952.

[1] J. Igusa, "On the Picard varieties attached to algebraic varieties," *American Journal of Mathematics*, vol. 74 (1952), pp. 1-22.

[2] A. Weil, "On Picard varieties," *American Journal of Mathematics*, vol. 74 (1952), pp. 865-894.

[3] A. Weil, *loc. cit.*, § III. Weil's proof of this theorem holds only for the case when the parameter variety is an analytic manifold (i. e. a variety without multiple points); however, it can be extended to the general case of an arbitrary analytic parameter variety by using some of the concepts and results developed in our (unpublished) paper cited in footnote 10.

duality between Abelian varieties introduced by Weil.[4] Our theory holds only for an algebraic variety and is hence possibly less general than that of Weil; as a compensation, we are able to formulate and prove the analyticity of the mapping in a somewhat stronger sense than that given by Weil's theory.[5] In the last section we shall deal with the more algebraic aspects of the problem; on the basis of some results on the Jacobian varieties proved by us elsewhere,[6] we investigate the fields of definition of some of the varieties and transformations introduced in the previous section. These results enable us to construct in a very simple way a system of divisors which represents the group $\mathfrak{G}_0(V)/\mathfrak{G}_l(V)$ in a "generally" one-to-one manner. This is however a weaker result than that obtained by Weil,[7] who has constructed a system of divisors which represents $\mathfrak{G}_0(V)/\mathfrak{G}_l(V)$ in an exactly one-to-one manner. Finally, we should like to mention that the method used in this last section has far greater potentialities than the immediate objective there indicates; in fact, this method is part of an idea which can be developed into a theory of Picard varieties for algebraic varieties over an arbitrary ground field, as will be shown in a forthcoming paper of ours on this subject.

We add here a few remarks concerning terminology. In this paper, the expression "analytic" shall always mean complex-analytic; and, unless explicitly stated to the contrary, the dimension is always the complex dimension, which is twice the topological dimension. By a cycle is always meant an algebraic cycle, except when explicitly stated to be a topological cycle, in which case the dimension is of course understood to be topological. When we are dealing with mappings of group varieties into one another, the expressions homomorphism, isomorphism, automorphism, etc. shall always mean analytic or algebraic homomorphism, isomorphism, automorphism, etc., according as we are dealing with analytic or algebraic varieties. The possibility of any confusion on this score is excluded by the theorem [8] that any (closed) analytic subvariety of an algebraic variety or analytic correspondence of an algebraic variety into another is an algebraic subvariety or an algebraic correspondence respectively.

---

[4] A. Weil, *loc. cit.*, § I.

[5] We shall prove that the mapping is holomorphic in the sense defined in section 2, while Weil's method (extended as indicated in footnote 3) yields only the somewhat weaker result that the mapping is analytic.

[6] W. L. Chow, "The Jacobian variety of an algebraic curve," to appear in print soon.

[7] A. Weil, *loc. cit.*, § II.

[8] W. L. Chow, "On compact complex analytic varieties," *American Journal of Mathematics*, vol. 71 (1949), pp. 893-914, Theorem V and Theorem VII.

2. Analytic mapping of cycles. By means of the method of associated forms [9] the aggregate of all positive cycles of a given dimension d and a given degree m (i. e. the degree in the ambient projective space) in an algebraic variety V can be represented in a one-to-one and algebraic manner by the points of an (in general reducible) algebraic variety, which we shall denote by $M(V; d, m)$. In fact, the (homogeneous) coordinates of a point in $M(V; d, m)$ can be taken to be the coefficients of the associated form of the positive cycle corresponding to it. We shall call this representation the canonical mappings of the positive cycles into $M(V; d, m)$. The canonical mapping has the property that if an analytic system of positive cycles $|Z|$ of dimension d and degree m in V is parametrized by an analytic variety W, i. e. if there is a closed analytic correspondence [10] between W and V such that to each point in W corresponds (with a suitable definition of multiplicity) a positive cycle of dimension d and degree m, then this parametrization induces an everywhere single-valued analytic transformation of W into $M(V; d, m)$. In case this last transformation is a holomorphic transformation of W into V, then $|Z|$ is said to be regularly parametrized by W. It is easily seen that if W is a manifold (i. e. a variety without multiple points), then the parametrization is always regular. If $|Z_1|$ and $|Z_2|$ are two analytic systems of positive cycles of the same dimension d and of degrees m_1 and m_2 respectively, both parametrized by the same analytic variety W, then the difference system $|Z_1 - Z_2|$, consisting of all

[9] W. L. Chow and B. L. van der Waerden, "Zur algebraischen Geometrie. IX. Ueber zugeordnete Formen und algebraische Systeme von algebraischen Mannigfaltigkeiten," *Mathematische Annalen*, Band 113 (1937), pp. 692-704; see also B. L. van der Waerden, *Einfuehrung in die algebraische Geometrie*, §§ 36-7, pp. 153-162. More precisely, it is proved there that there is an algebraic correspondence (which is in general reducible) between $M(V; d, m)$ and V such that to each point of $M(V; d, m)$ there corresponds under this correspondence a positive cycle of dimension d and degree m in V, the multiplicities of the components of the cycle being determined by the associated form corresponding to the point in $M(V; d, m)$, and that in this way all positive cycles of dimension d and degree m in V can be obtained. This is what is meant by the statement that the positive cycles of dimension d and degree m in V are represented by the points of $M(V; d, m)$ in a one-to-one and algebraic manner.

[10] We shall use in this section some of the concepts developed in our paper, "On complex analytic varieties," to appear soon in this Journal. We refer to that paper for the definition of a complex analytic variety in abstracto and the definition of a holomorphic function on such a variety. We mention here that a closed analytic correspondence of an analytic variety W into an analytic variety V is defined by a closed analytic subvariety in the product variety $W \times V$, and that such a correspondence is called an everywhere single-valued analytic transformation, if to each point in W corresponds exactly one point in V. A single-valued analytic transformation of W into V is said to be a holomorphic transformation, if every holomorphic function on V is carried by the inverse transformation into a holomorphic function on W.

cycles $Z_1 - Z_2$ as the common parameter runs through the variety W, is said to be an analytic system of cycles of dimension d and degree $m = m_1 - m_2$, and W is its parameter variety. The parametrization of $| Z_1 - Z_2 |$ by W is said to be regular, if both $| Z_1 |$ and $| Z_2 |$ are regularly parametrized by W.

Suppose we have a mapping $f(Z)$, which assign to each element of an aggregate $\{Z\}$ of cycles of a given dimension d in V a point in an analytic variety U. We shall say that this mapping $f(Z)$ is analytic, if for any analytic system of cycles in $\{Z\}$ parametrized by an analytic variety W, the mapping $f(Z)$ induces an (everywhere single-valued) analytic transformation of W into U. The mapping $f(Z)$ is said to be holomorphic (or regularly analytic) if this induced transformation is holomorphic whenever W is a regular parametrization of the analytic system. Assume that U is an Abelian variety and that the mapping $f(Z)$ is a homomorphism of a subgroup $\{Z\}$ of the group of all cycles of dimension d in V into U; then the mapping $f(Z)$ is essentially determined by the mapping of the positive cycles in $\{Z\}$. It is then easily seen that the mapping $f(Z)$ is holomorphic, if for the positive cycles of degree m in $\{Z\}$ (for any m) the mapping $f(Z)$ is the product of the canonical mapping of Z into $M(V; d, m)$ and a holomorphic transformation of a subvariety of $M(V; d, m)$ into U.

Let Y be an algebraic subvariety of dimension $r - d$ in V (r being the dimension of V) which is the intersection of V with a linear subspace L of the proper dimension in the ambient space, and let $\{Z\}$ be the aggregate of all positive cycles of dimension d and degree m in V which has a proper intersection with Y. Let $f(Z)$ be the mapping of $\{Z\}$ in $M(Y; 0, m)$ which is the product of the mapping $Z \to Z \cdot Y$ and the canonical mapping of $Z \cdot Y$ into $M(Y; 0, m)$; then $f(Z)$ is a holomorphic mapping. In fact, the mapping $Z \to Z \cdot Y$ induces a mapping f' of an open subset[11] in $M(V; d, m)$ into

[11] This subset is not only open in the sense of the ordinary topology, but also in the sense of the topology where the closed subsets are the (reducible or irreducible) algebraic subvarieties. To show this, we recall the fact that if $F(u^{(0)}, u^{(1)}, \cdots, u^{(d)})$ is the associated form of a positive cycle Z of dimension d and degree m and if $v^{(i)} = (v^{(i)}_0, v^{(i)}_1, \cdots, v^{(i)}_n)$, $i = 0, 1, \cdots, d$, are the hyperplane coordinates of any $d + 1$ hyperplanes in the ambient projective space, then we have $F(v^{(0)}, v^{(1)}, \cdots, v^{(d)}) = 0$ if and only if these $d + 1$ hyperplanes have a common intersection point with Z. See, e. g., B. L. van der Waerden, *loc. cit.*, p. 157. Now, if we let $v^{(1)}, \cdots, v^{(d)}$ be the hyperplane coordinates of any d hyperplane whose intersection is the linear subspace L, then $F(u^{(0)}, v^{(1)}, \cdots, v^{(d)})$ will vanish identically in the variables $u^{(0)}$ if and only if the cycle Z does not intersect L and hence also Y properly. The condition that $F(u^{(0)}, v^{(1)}, \cdots, v^{(d)})$ vanishes identically in the variables $u^{(0)}$ can obviously be expressed as a system of homogeneous algebraic equations in the coefficients of the form $F(u^{(0)}, u^{(1)}, \cdots, u^{(d)})$, which are the coordinates of the point in $M(V; d, m)$ representing the

$M(Y; 0, m)$; and since $f(Z)$ is evidently the product of the canonical mapping of Z into $M(V; d, m)$ and f', it is sufficient to show that this mapping f' is a holomorphic transformation. Now, the coordinates of the point in $M(V; d, m)$ representing a cycle Z are the coefficients of the associated form of Z, and the associated form of $Z \cdot Y$ can be obtained by substituting for certain sets of variables in the associated form of Z the coefficients of the linear equations defining the linear subspace L. Furthermore, by a well known property of the associated form mentioned in footnote 11, the so obtained associated form of $Z \cdot Y$ cannot vanish identically so long as Z intersects Y properly. As the coefficients of the associated form of $Z \cdot Y$ are the coordinates of the point in $M(Y; 0, m)$ representing $Z \cdot Y$, it follows that the mapping f' is a holomorphic transformation.

3. **Dual pairing of complex tori.** Following Weil we shall introduce the concept of duality between complex tori of a given dimension; a dual pair of complex tori is characterized by the properties that the one is the character group of the fundamental group of the other and that the automorphism of the covering group of the one corresponding to the multiplication by $\sqrt{-1}$ induces the same automorphism of the covering group of the other. We can describe this concept in non-invariant terms as follows. Let S be the complex vector space of dimension p, and let a coordinate system be chosen as that a vector in S is represented by a point $(z) = (z_1, \cdots, z_p)$ in the p-dimensional complex number space. Let $\mathfrak{P}$ be a $(2p, p)$-matrix whose rows represent $2p$ real-linearly independent vectors, and let $[\mathfrak{P}]$ be the subgroup of S generated by these $2p$ vectors. Then the quotient group $U = S/[\mathfrak{P}]$ is a complex torus of dimension p. Let S' be another complex vector space of dimension p, and let a coordinate system be chosen so that a vector in S' is also represented by a point $(z') = (z'_1, \cdots, z'_p)$ in the p-dimensional complex number space. We shall consider S' to be invariantly associated with S in such a way that if S undergoes a coordinate transformation $(z) \rightarrow (z)\mathfrak{T}$, then S' undergoes the coordinate transformation $(z') \rightarrow (z')\,{}^t\bar{\mathfrak{T}}^{-1}$. Let $\mathfrak{P}'$ be the $(2p, p)$-matrix such that $(\mathfrak{P}'\bar{\mathfrak{P}}')\,{}^t(\mathfrak{P}\bar{\mathfrak{P}}) = 1$; then the quotient group $U' = S'/[\mathfrak{P}']$ is a complex torus of dimension p which is invariantly associated with U and which can be considered as the character group of $[\mathfrak{P}]$ by assigning to each point in U', represented by a point (z') mod. $[\mathfrak{P}']$, the character

$$\tfrac{1}{2}(\bar{z}')\,{}^t(z) + \tfrac{1}{2}(z')\,{}^t(\bar{z}) \pmod{1} \tag{1}$$

cycle Z; it follows then that the set of all points in $M(V; d, m)$ which represents cycles having improper intersections with Y is an (in general reducible) algebraic subvariety in $M(V; d, m)$.

for all points (z) in $[\mathfrak{P}]$. It is clear that this relation between U and U' is mutual; they are called the dual of each other, and we shall say that they are dually paired by the form (1).

The duality between U and U' can be extended to the (closed, connected) analytic subgroups and the corresponding factor groups, which are also all complex tori. Let U_0 be an analytic subgroup of dimension $p-q$ in U; then the component of the trace of U_0 in S which contains the zero vector is also an analytic subgroup and hence must be a linear subspace S_0 of dimension $p-q$, so that we have $U_0 = S_0/S_0 \cap [\mathfrak{P}]$ and hence U_0 is a complex torus. Without any loss of generality, we can assume the coordinate system of S so chosen that S_0 is the subspace defined by $z_1 = \cdots = z_q = 0$; let S_1 be the subspace defined by $z_{q+1} = \cdots = z_p = 0$. Similarly, let S'_0 be the subspace of S' defined by $z'_1 = \cdots = z'_q = 0$, and let S'_1 be the subspace defined by $z'_{q+1} = \cdots = z'_p = 0$. The fact that $S_0/S_0 \cap [\mathfrak{P}]$ is a compact manifold implies that the group $S_0 \cap [\mathfrak{P}]$ contains $2p-2q$ real-linearly independent vectors; hence, by a well known property of Abelian groups, it must possess a base which can be extended to a base of $[\mathfrak{P}]$. By a change of base of $[\mathfrak{P}]$ if necessary (which corresponds to the multiplication of $\mathfrak{P}$ by a unimodular integral matrix on the left), we can therefore assume that the last $2p-2q$ rows of $\mathfrak{P}$ constitute a base of $S_0 \cap [\mathfrak{P}]$. The matrix $\mathfrak{P}$ will then have the form $\begin{pmatrix} \mathfrak{P}_1 & * \\ 0 & \mathfrak{P}_0 \end{pmatrix}$, so that we have $S_0 \cap [\mathfrak{P}] = [\mathfrak{P}_0]$, and the corresponding matrix $\mathfrak{P}'$ will have the form $\begin{pmatrix} \mathfrak{P}'_1 & 0 \\ * & \mathfrak{P}'_0 \end{pmatrix}$. The relation $(\bar{\mathfrak{P}}'\mathfrak{P}')^t\mathfrak{P}\bar{\mathfrak{P}}) = 1$ implies then the relations $\bar{\mathfrak{P}}'_1{}^t\mathfrak{P}_1 + \mathfrak{P}'_1{}^t\bar{\mathfrak{P}}_1 = 1$ and $\bar{\mathfrak{P}}'_0{}^t\mathfrak{P}_0 + \mathfrak{P}'_0{}^t\bar{\mathfrak{P}}_0 = 1$; this shows that the rows of each of the matrices $\mathfrak{P}_1$, $\mathfrak{P}'_0$, $\mathfrak{P}'_1$ are also real-linearly independent and hence the groups $U_1 = S_1/[\mathfrak{P}_1]$, $U'_0 = S'_0/[\mathfrak{P}'_0]$, $U'_1 = S'_1/[\mathfrak{P}'_1]$ are all complex tori (of dimensions q, $q-p$, q respectively). Furthermore, these relations show that both U_0, U'_0 and U_1, U'_1 can be made into dual pairs, if we define the character relations in each pair by the same form (1) as restricted to the pairs of subspaces S_0, S'_0 and S_1, S'_1 respectively. The fact that the rows of $\mathfrak{P}'_0$ are real-linearly independent shows that $S'_1 \cap [\mathfrak{P}'] = [\mathfrak{P}'_1]$, and hence $U'_1 = S'_1/S'_1 \cap [\mathfrak{P}']$ can be considered as an analytic subgroup of U'. It is also clear that U_1 is isomorphic to the factor group U/U_0 and U'_0 is isomorphic to the factor group U'/U'_1, so that we can set $U_1 = U/U_0$ and $U'_0 = U'/U'_1$ and thus obtain the dual pairs U_0, U'/U'_1 and U/U_0, U'_1. Finally, we observe that U'_1, considered as a subgroup of the character group U' of $[\mathfrak{P}]$, is precisely the annihilator of $[\mathfrak{P}_0]$, and hence can be considered as the character group of $[\mathfrak{P}]/[\mathfrak{P}_0]$ under the canonical

homomorphism of $[\mathfrak{P}]$ onto $[\mathfrak{P}]/[\mathfrak{P}_0]$; and that if we identify $[\mathfrak{P}]/[\mathfrak{P}_0]$ with $[\mathfrak{P}_1]$ by projecting S onto the subspace S_1, then for each element in U'_1 represented by a point (z') mod. $[\mathfrak{P}'_1]$, this induced character is defined by the form (1) for all (z) in $[\mathfrak{P}_1]$. Thus the dual pairing of U/U_0 and U'_1 is an analytic extension of the well known group-theoretic pairing of $[\mathfrak{P}]/[\mathfrak{P}_0]$ and U'_1, and we shall express this fact by saying that the dual pairing of U/U_0 and U'_1 is subordinate to the given dual pairing of U and U'. Similarly, the dual pairing of U_0 and U'/U'_1 if of course also subordinate to the dual pairing of U and U'.

A complex torus U is said to be an Abelian variety, if U is an algebraic variety, i. e., more precisely, if there is a one-to-one bi-holomorphic transformation of U onto an algebraic variety in some projective space. It follows from this definition and the theorem [8] quoted in section 1 that any analytic subgroup U_0 in an Abelian variety U is also an Abelian variety. Furthermore, it is known from the theory of Riemann matrices that the dual of an Abelian variety is also an Abelian variety; it follows then from this and the above paragraph that also the factor group U/U_0 is an Abelian variety.[12]

4. The Jacobian varieties. We shall give in this section a brief resumé of some properties of the Jacobian varieties which will be useful to us later. Let C be a non-singular algebraic curve of genus p. Let $(\alpha) = (\alpha_1, \cdots, \alpha_{2p})$ be a base of its first Betti group $\mathfrak{B}(C)$, and let I_α be the intersection matrix of this base. Let $(\int\Phi) = (\int\Phi_1, \cdots, \int\Phi_p)$ be a system of p independent Abelian integrals of the first kind on C, and let Ω be the $(2p, p)$-matrix whose i-th row is the vector $(\int_{\alpha_i}\Phi_1, \cdots, \int_{\alpha_i}\Phi_p)$. The complex torus $J = S/[\Omega]$ is called the Jacobian variety of C, and the dual complex torus $J' = S'/[\Omega']$ is called the dual Jacobian variety of C; it is known that J is an Abelian variety and that J' is isomorphic to J and hence is also an Abelian variety. In fact, if we set $\mathfrak{R} = {}^t\bar{\Omega}I_\alpha^{-1}\Omega$, then we will have from the known properties

[12] It can be shown by the method of associated forms that the factor group U/U_0 of an Abelian variety U over an Abelian subvariety (or a finite subgroup) U_0 is also an Abelian variety. More precisely, we have the following theorem: Let T be a homomorphism of an Abelian variety U into an Abelian variety U', and let K be a field over which U, U' and T are all defined; then the kernel U_0 of T is a rational cycle over K and the factor group U/U_0 is an Abelian variety which can be represented by an algebraic variety defined over K; furthermore, the homomorphism T induces then an isomorphism of the so represented Abelian variety U/U_0 into U', which is also defined over K. The proof of this theorem will be given in our forthcoming paper mentioned in section 1.

of the Riemann matrix Ω the relation $I_\alpha^{-1}\Omega = \Omega'\mathfrak{N}$; this shows that the linear transformation $(z) = (z')\mathfrak{N}$ of S' onto S induces an isomorphism of J' onto J, which we shall denote by N.

Let x_0 be any fixed point in C; then for any point x in C the mapping

$$f(x) = \left(\int_{x_0}^{x_1}\Phi_1,\cdot\cdot\cdot,\int_{x_0}^{x_1}\Phi_p\right) \qquad (\text{mod. } [\Omega])$$

is a holomorphic transformation of C into J, which can be extended to any 0-cycle $Z = \Sigma\, n_i x_i$ in C by the stipulation $f(Z) = \Sigma\, n_i f(x_i)$. This mapping $f(Z)$ is holomorphic. To show this we can without any loss of generality restrict ourselves to the case where Z is a positive 0-cycle of a given degree m in C. Then the corresponding variety $M(C;0,m)$ is simply the m-th symmetric product of the curve C and hence is a non-singular algebraic variety.[13] The mapping $f(Z)$ induces an everywhere single-valued analytic transformation of $M(C;0,m)$ into J, which on account of the non-singular nature of $M(C;0,m)$ must be a holomorphic transformation. As $f(Z)$ is evidently the product of the canonical mapping of Z into $M(C;0,m)$ and this last holomorphic transformation, our assertion is therefore proved.

Let $\int\Psi$ be the Abelian integral of the third kind on C which has the 0-cycle $x - x_0$ as its residue divisor and whose periods are all pure imaginary. We can then define a character of $\mathfrak{B}(C)$ by assigning to each element α in $\mathfrak{B}(C)$ the real number $\dfrac{1}{2\pi\sqrt{-1}}\displaystyle\int_\alpha \Psi(\text{mod. } 1)$. If we identify $\mathfrak{B}(C)$ with $[\Omega]$ by means of the isomorphic mapping $\sum_{i=1}^{2p} m_i\alpha_i \to (m)\Omega$, then this character is represented by a point $g(x)$ in J'. In fact, we have the formula

$$g(x) = \frac{1}{2\pi\sqrt{-1}}\left(\int_{\alpha_1}\Psi,\cdot\cdot\cdot,\int_{\alpha_{2p}}\Psi\right)\Omega' \qquad (\text{mod. } [\Omega']),$$

and this mapping $g(x)$ of C into J' can be extended to any 0-cycle $Z = \Sigma\, n_i x_i$ in C by the stipulation $g(Z) = \Sigma\, n_i g(x_i)$. It is clear that $g(Z) = 0$ (mod. $[\Omega']$) if and only if Z is a principal divisor, so that $g(Z)$ induces an isomorphic mapping of $\mathfrak{G}_c(C)/\mathfrak{G}_l(C)$ into J'. Furthermore, this mapping $g(Z)$ is also holomorphic. In fact, let $\int\phi_1,\cdot\cdot\cdot,\int\phi_{2p}$ be the harmonic integrals on C corresponding to the topological 1-cycles $\alpha_1,\cdot\cdot\cdot,\alpha_{2p}$ respectively, and let $\mathfrak{M}$ be a $(p,2p)$-matrix such that

$$\left(\int\phi_1,\cdot\cdot\cdot,\int\phi_{2p}\right) = \left(\int\bar{\Phi}_1,\cdot\cdot\cdot,\int\bar{\Phi}_p,\int\Phi_1,\cdot\cdot\cdot,\int\Phi_p\right)\begin{pmatrix}\bar{\mathfrak{M}}\\ \mathfrak{M}\end{pmatrix}.$$

---

[13] See B. L. van der Waerden, "Divisorklassen in algebraischen Funktionenkoerpern," *Commentarii Mathematici Helvetici*, vol. 20 (1947), pp. 68-109, § 13.

Then we have

$$\begin{aligned} g(x) &= \left(\int_{x_0}^{x} \phi_1, \cdots, \int_{x_0}^{x} \phi_{2p} \right) \Omega' \\ &= \left(\int_{x_0}^{x} \bar{\Phi}_1, \cdots, \int_{x_0}^{x} \bar{\Phi}_p, \int_{x_0}^{x} \Phi_1, \cdots, \int_{x_0}^{x} \Phi_p \right) \begin{pmatrix} \bar{\mathfrak{M}} \\ \mathfrak{M} \end{pmatrix} \Omega' \\ &= \left(\int_{x_0}^{x} \bar{\Phi}_1, \cdots, \int_{x_0}^{x} \bar{\Phi}_p, \int_{x_0}^{x} \Phi_1, \cdots, \int_{x_0}^{x} \Phi_p \right) {}^t(\Omega' \bar{\Omega}') I_\alpha \Omega' \\ &= \left(\int_{x_0}^{x} \Phi_1, \cdots, \int_{x_0}^{x} \Phi_p \right) \mathfrak{N}^{-1} \qquad \qquad (\text{mod. } [\Omega']) ; \end{aligned}$$

this shows that $g(x) = N^{-1}f(x)$, which proves our assertion. Thus in the isomorphism N of J' onto J, the mapping $g(x)$ is carried over into the mapping $f(x)$; this shows that these two mappings are essentially the same, being given here two different interpretations as mappings into a pair of dual Abelian varieties.

5. The Picard varieties. Let V be a non-singular algebraic variety of dimension r embedded in a projective space, and let C be a non-singular algebraic curve in V such that any topological 1-cycle in V is homologous to a topological 1-cycle in C. Such a curve C certainly exists; we can take, for example, the intersection of V with a sufficiently general linear subspace of the complementary dimension in the ambient space of V.[14] If $2q$ is the rank of the first Betti group $\mathcal{B}(V)$ of V, then there exist (according to a theorem of Hodge) exactly q independent Abelian integrals of the first kind $\int \Theta_1, \cdots, \int \Theta_q$ in V. Let $(\gamma) = (\gamma_1, \cdots, \gamma_{2q})$ be a base of $B(V)$, and let Ω_1 be the $(2q, q)$-matrix whose i-th row is the vector $\left(\int_{\gamma_i} \Theta_1, \cdots, \int_{\gamma_i} \Theta_q \right)$. Since the rows of Ω_1 are real-linearly independent, the quotient group $S_1/[\Omega_1]$ is a complex torus of dimension q, which we shall call the Albanese variety A attached to V. The dual complex torus P of A is then called the Picard variety attached to V. Let x_0 be any fixed point in V, which we shall assume for the sake of convenience to be also a point in C and take as the reference

[14] That such an intersection has the indicated property is usually given a very sketchy proof in the literature. For a more detailed treatment, see W. L. Chow, "On the fundamental group of an algebraic variety," *American Journal of Mathematics*, vol. 74 (1952), pp. 726-736, Theorem 1.

point in the definition of the mappings $f(x)$ and $g(x)$ in the previous section; then, for any point x in V the mapping

$$F(x) = \left(\int_{x_0}^{x}\Theta_1,\cdots,\int_{x_0}^{x}\Theta_q\right) \qquad (\mathrm{mod.}\ [\Omega_1])$$

is a holomorphic transformation of V into A, which can be extended to any 0-cycle $Z = \Sigma\, n_i x_i$ in V by the stipulation $F(Z) = \Sigma\, n_i F(x_i)$. It is easily seen that $F(Z)$ is an analytic mapping of the 0-cycles in V into A; however, in contrast to the case of a curve, we do not know whether this mapping is also holomorphic.

The inclusion mapping of C into V induces a homomorphism L of $\mathfrak{B}(C)$ into $\mathfrak{B}(V)$, which can be expressed by the relation $(\alpha) \approx (\gamma)\mathfrak{L}$ on V, where $\mathfrak{L}$ is an integral $(2q, 2p)$-matrix. The assumption that every topological 1-cycle in V is homologous to a topological 1-cycle in C implies that this homomorphism L is onto, i. e. there exixsts an integral $(2p, 2q)$-matrix $\mathfrak{L}'$ such that $\mathfrak{L}\mathfrak{L}' = 1$. Since the integrals $\int\Theta_1,\cdots,\int\Theta_q$ are also Abelian integrals of the first kind on the curve C, there exists a (p, q)-matrix $\mathfrak{T}$ such that the relation $(\int\Theta) = (\int\Phi)\mathfrak{T}$ holds on the curve C. Integrating both sides over the base (α), we obtain the equation ${}^t\mathfrak{L}\Omega_1 = \Omega\mathfrak{T}$ and hence the equation $\Omega_1 = {}^t\mathfrak{L}'\Omega\mathfrak{T}$; we have then the equation $[\Omega\mathfrak{T}] = [\Omega_1]$ and hence the mapping $(z) \to (z)\mathfrak{T}$ induces a homomorphism T of J onto A. The kernel J_0 of T is an analytic subgroup of dimension $p - q$ in J; furthermore, if we denote by S_0 the linear subspace of dimension $p - q$ in S defined by the equation $(z)\mathfrak{T} = 0$, then we have evidently $J_0 = S_0/S_0 \cap [\Omega]$ and hence J_0 is an Abelian variety. The homomorphism T induces an isomorphism of the factor group J/J_0 onto A, which we shall also denote by T; since J/J_0 is an Abelian variety, it follows that A and consequently also P are Abelian varieties. Let the subgroup J'_1 in J' be the annihilator of $S_0 \cap [\Omega]$; we have shown before that J'_1 is the dual of J/J_0 and that the dual pairing of J/J_0 and J'_1 is subordinate to the dual pairing of J and J'. The isomorphism T of J/J_0 onto A then induces an isomorphism T' of P onto J'_1, such that the dual pairing of A and P is carried by these isomorphisms into the dual pairing of J/J_0 and J'_1.

Let D be a divisor in V which is homologous to zero. According to the Weil's generalization of a theorem of Lefschetz,[15] there exists an Abelian integral $\int\Psi$ of the third kind on V with D as its residue divisor. Adding a linear combination of the Abelian integrals of the first kind if necessary,

[15] A. Weil, "Sur la théorie des formes différentielles attachées à une variété analytique complexe," *Commentarii Mathematici Helvetici*, vol. 20 (1947), pp. 110-116.

we can assume that the periods of $\int\Psi$ are all pure imaginary; then the integral $\int\Psi$ is uniquely determined by the divisor D up to an additive constant. We can then define a character of $\mathfrak{B}(V)$ by assigning to each element γ in $\mathfrak{B}(V)$ the real number $\frac{1}{2\pi\sqrt{-1}}\int_\gamma \Psi \pmod{1}$. If we identify $\mathfrak{B}(V)$ with $[\Omega_1]$ by means of the isomorphic mapping $\sum_{i=1}^{2q} m_i\gamma_i \to (m)\Omega_1$, then this character is represented by a point $G(D)$ in P. Thus we have a homomorphism $G(D)$ of the group $\mathfrak{G}_h(V)$ of all divisors in V which are homologous to zero, into the Abelian variety P. By considering the multiplicative function $\exp(\int_{x_0}^{x}\Psi)$ on V, it can be easily seen that the kernel of this homomorphism is precisely the group $\mathfrak{G}_l(V)$, so that $G(D)$ induces an isomorphism of $\mathfrak{G}_h(V)/\mathfrak{G}_l(V)$ into P. We shall show that this mapping $G(D)$ is holomorphic. We observe first that if the intersection 0-cycle $C\cdot D$ is defined, then the integral $\int\Psi$ can also be considered as an integral of the third kind on the curve C with $C\cdot D$ as its residue; and as such it defines a character of $\mathfrak{B}(C)$ which is the same as that induced by the corresponding character of $\mathfrak{B}(V)$ under the homomorphism L of $\mathfrak{B}(C)$ onto $\mathfrak{B}(V)$. We have then the equation $G(D)=T'^{-1}g(C\cdot D)$; we observe here that the mapping $T'^{-1}g(C\cdot D)$ is defined for any divisor D (not necessarily homologous to zero) such that $C\cdot D$ is defined. Now, let $D(w)=D_1(w)-D_2(w)$ be an analytic system of divisors in V, (regularly) parametrized by an analytic variety W, where $D_1(w)$ and $D_2(w)$ are positive divisors such that $D_1(w)\sim D_2(w)$ for all w. If w_0 is any point in W, we can choose the curve C such that the intersection 0-cycle $C\cdot D(w_0)$ is defined; then the intersection 0-cycle $C\cdot D(w)$ is defined for all points w in a sufficiently small neighborhood of w_0. Furthermore, we can assume that C is the intersection of V with a suitably chosen linear subspace in the ambient space. Then we have the equation

$$G(D(w))=T'^{-1}g(C\cdot D(w))=T'^{-1}g(C\cdot D_1(w))-T'^{-1}g(C\cdot D_2(w))$$

for all points w in a neighborhood of w_0. The mapping $T'^{-1}g(C\cdot D_1(w))$ is obviously the product of the canonical mapping of $D_1(w)$ into $M(V;r-1,m)$ (m being the degree of $D_1(w)$), the holomorphic transformation of $M(V;r-1,m)$ into $M(C;0,m)$ induced by the mapping $D\to C\cdot D$ (see the last paragraph of section 1), the holomorphic transformation of $M(C;0,m)$ into J' induced by $g(Z)$, and the isomorphism T'^{-1} of J'_1 onto P; it follows then that $T'^{-1}g(C\cdot D_1(w))$ is an analytic (holomorphic) transformation of a neighborhood of w_0 into P, and similarly also $T'^{-1}g(C\cdot D_2(w))$

is an analytic (holomorphic) transformation. Thus we have shown that $G(D(w))$ is an analytic (holomorphic) transformation of a neighborhood of w_0 into P; and since w_0 is any point in W, it follows that $G(D(w))$ is an analytic (holomorphic) transformation of W into P. This proves our assertion.

In order to complete the proof of Igusa's duality theorem, it only remains to show that $\mathfrak{G}_h(V) = \mathfrak{G}_a(V)$ and that the mapping $G(D)$ of $\mathfrak{G}_h(V)$ into P is onto. Since $\mathfrak{G}_a(V)$ is a subgroup of $\mathfrak{G}_h(V)$, both assertions will follow if we show that $G(D)$ maps the group $\mathfrak{G}_a(V)$ onto P; and this can be done by the construction of the so-called Poincaré family of divisors. However, we shall construct in the next section an analytic system of divisors which represents the group $\mathfrak{G}_a(V)/\mathfrak{G}_l(V)$ in a more precise manner.

6. Representative system of divisors. We begin by observing (what we have implicitly assumed up to now) that the field $\mathfrak{K}$ of all complex numbers is the universal domain, in the sense of Weil,[16] of the algebraic geometry we are concerned with here. We shall assume that the Abelian variety A is represented by an arbitrarily chosen, but fixed non-singular algebraic variety in some projective space; it follows then that the mapping $F(x)$ of V into A is a rational transformation, defined (in the sense of Weil) for all points of V. Let K be a field (i. e., in the terminology of Weil, a subfield of $\mathfrak{K}$ such that $\mathfrak{K}$ has infinite degree of transcendency over K) over which the algebraic varieties V and A, as well as the rational transformation $F(x)$ and the group composition in A, are all defined; furthermore, we shall assume that the point x_0, the reference point in the definition of $F(x)$, is rational over K. We can, in fact, take K to be a finitely generated extension of the rational field. Consider an $(r-1)$-dimensional linear system $|C|$ of curves on V, which is cut out on V by a system of linear subspaces in the ambient space; by choosing the system of linear subspaces in a suitable manner we can assume that the system $|C|$ has the following properties: (1) the system $|C|$ is defined over K, i. e. the associated variety W of $|C|$ in $M(V; 1, m)$, m being the degree of $|C|$, is defined over K; for any point w in W, we shall denote by $C(w)$ the curve of the system corresponding to it; (2) the point x_0 is a base point of $|C|$, and a generic point of V over K is contained in exactly one curve of the system $|C|$; (3) for any generic point u of W over K, the curve $C(u)$ is a non-singular curve, whose genus

[16] A. Weil, *Foundations of Algebraic Geometry*, American Mathematical Society Colloguium Publication no. 29 (1946). We shall use in this section the algebraic-geometric concepts and terminology developed in this book.

will be denoted by p. According to the theorem cited in footnote 14, the curve $C(u)$ will have the property (stated at the beginning of section 5) that every topological 1-cycle in V is homologous to a topological 1-cycle in $C(u)$. We can therefore apply the results of sections 4 and 5 to the curve $C(u)$; let $J(u)$ be the Jacobian variety of $C(u)$ and let T_u be the homomorphism of $J(u)$ onto A defined in section 5. Since $C(u)$ is defined over $K(u)$, it follows from a result proved elsewhere [17] that we can consider $J(u)$ as represented by a non-singular algebraic variety, such that both the variety $J(u)$ as well as the group composition in $J(u)$ are defined over $K(u)$. Moreover, since the point x_0 is a rational divisor of degree 1 in $C(u)$ over $K(u)$, the "canonical mapping" f_u of $C(u)$ into $J(u)$, defined with x_0 as the reference point, is also a rational transformation defined over $K(u)$. Then the homomorphism T_u is a rational transformation of $J(u)$ onto A; we maintain that it is also defined over $K(u)$. In fact, the rational transformation $F(x)$ of V into A induces a rational transformation of $C(u)$ into A, which is evidently defined over $K(u)$. Since A is an Abelian variety, this rational transformation has a uniquely determined linear extension,[18] which is a rational transformation of $J(u)$ into A, also defined over $K(u)$; and it can be easily seen, in view of our choice of the reference point x_0 for both $F(x)$ and f_u, that this rational transformation is precisely the homomorphism T_u.

The fact that T_u is defined over $K(u)$ is the basis for the construction of what is usually called the Poincaré family of divisors,[19] which is "generically" parametrized by A. Though this family does contain a representative for each element in $\mathfrak{G}_a(V)/\mathfrak{G}_l(V)$, it has the disadvantage that this representative is in general not uniquely determined, even for a "generic" element in $\mathfrak{G}_a(V)/\mathfrak{G}_l(V)$, there being a finite number of them in the family; this is due to the fact that A can be represented as a factor group $J(u)/J_0(u)$ of $J(u)$, but not necessarily as a subgroup of $J(u)$. In order to get a "generally" one-to-one representative system of divisors for $\mathfrak{G}_a(V)/\mathfrak{G}_l(V)$, we shall show that the isomorphism T'_u of P into $J'(u)$, as defined in section 5, is also defined over $K(u)$, provided we represent both P and $J'(u)$ by suitably chosen algebraic varieties. In fact, since $J(u)$ and $J'(u)$ are

---

[17] See the paper cited in footnote 6.

[18] See A. Weil, *Variétés abéliennes et courbes algébriques*, p. 77, Théorème 21.

[19] We have in mind here the method of construction due to F. Enriques, assuming the existence of q independent Abelian integrals of the first kind in V. See O. Zariski, *Algebraic Surfaces*, Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 3, no. 5 (1935), p. 123 and the references cited there.

isomorphic, we can represent both of them by the same algebraic variety and thus identify $J'(u)$ with $J(u)$ by means of the isomorphism N defined in section 4; we can then consider $J'_1(u)$ as an Abelian subvariety of $J(u)$ and we have evidently $g_u = f_u$. The construction of the algebraic variety representing P will be given in course of the proof below; we shall see that the so represented variety P is defined over K.[20]

Let X be a cycle of dimension q in $J(u)$, rational over $K(u)$; then for a generic point v of A over $K(u)$, the intersection $X \cdot T_u^{-1}(v)$ is a rational 0-cycle over $K(u, v)$. If for any 0-cycle Z in an Abelian variety we denote (with Weil[21]) by $\mathfrak{S}(Z)$ the element in the Abelian variety which is the sum of all points in Z, each point being added or subtracted as often as its (positive or negative) multiplicity indicates, then the point $\mathfrak{S}(X \cdot T_u^{-1}(v))$ in $J(u)$ is rational over $K(u, v)$; hence, the mapping $v \rightarrow \mathfrak{S}(X \cdot T_u^{-1}(v))$ is a rational transformation of A into $J(u)$, defined over $K(u)$. The mapping $H_u : v \rightarrow \mathfrak{S}(X \cdot T_u^{-1}(v)) - \mathfrak{S}(X \cdot T_u^{-1}(0))$, is then a homomorphism of A into $J(u)$, defined over $K(u)$, and hence the image $H_u(A)$ is an Abelian subvariety in $J(u)$, also defined over $K(u)$. It is clear that if s is the degree of the 0-cycle $X \cdot T_u^{-1}(v)$, then we have the equation $T_u H_u(v) = \mathfrak{S}(sv)$; this shows that $H_u(A)$ has the dimension q, and that the kernel A_0 of H_u is a finite subgroup of A and hence is a positive 0-cycle, rational over $K(u)$. Since the Abelian variety A is defined over K, the positive 0-cycle A_0, being a finite subgroup in A, must be algebraic over K; and as $K(u)$ is a regular extension of K, it follows that A_0 is rational over $K(u) \cap \bar{K} = K$. According to the theorem cited in footnote 12, the factor group A/A_0 is then an Abelian variety which can be represented by a non-singular algebraic variety, defined over K, and the homomorphism H_u of A into $J(u)$ induces an isomorphism T'_u of A/A_0 into $J(u)$, which is also defined over $K(u)$. We shall now show that $T'_u(A/A_0) = J'_1(u)$, so that A/A_0 is isomorphic to P and hence can be taken as a representative algebraic variety of P. In order to prove this, we only need to show that for a generic point y of A/A_0 over $K(u)$, there exists a divisor $D(y)$ in $G_0(V)$ such that $f_u(D(y) \cdot C(u)) = T'_u(y)$. We recall that $M(C; 0, p)$ is a non-singular (irreducible) algebraic variety of dimension p, defined over K; in fact, it is the symmetric p-th product of the curve $C(u)$.

---

[20] The first part of the proof below, namely the construction of the Abelian variety $H_u(A)$, is essentially the same as Castelnuovo's proof for Poincaré's "complete reducibility theorem." See G. Castelnuovo, "Sugli integrali semplici appartenenti ad una superficie irregolare," *Rend. della R. Accad. dei Lincei*, ser. V, vol. XIV (1905); compare also A. Weil, *loc. cit.*, p. 94, Théorème 26.

[21] A. Weil, *loc. cit.*, p. 28.

The mapping f_u of the cycles in $C(u)$ into $J(u)$ induces a birational transformation h_u of $M(C;0,p)$ onto $J(u)$, defined over $K(u)$; hence, since $T'_u(y)$ is a generic point of $T'_u(A/A_0)$ over $K(u)$, the inverse image $h_u^{-1}(T'_u(y))$ is a rational point of $M(C;0,p)$ over $K(u,y)$ and represents a positive 0-cycle $Z(u,y)$ of degree p in $C(u)$, which is rational over $K(u,y)$. Over the field $K(y)$, the mapping $u \rightarrow Z(u,y)$ defines an algebraic correspondence of W into V, and it is easily seen that the projection of this correspondence in V is a positive divisor $Z(y)$ in V, rational over $K(y)$, whose intersection with $C(u)$ is precisely $Z(u,y)$, apart from a possible fixed component Z_0. Let $Z(0)$ be a specialization of $Z(y)$ over the specialization $y \rightarrow 0$ over K, such that $C(u) \cdot Z(0) = px_0 + Z_0$; if K' is an extension of K over which the divisor $Z(0)$ is rational, then $D(y) = Z(y) - Z(0)$ is a divisor in $\mathfrak{G}_0(V)$ such that

$$\begin{aligned} f_u(D(y) \cdot C(u)) &= f_u(Z(y) \cdot C(u)) - f_u(Z(0) \cdot C(u)) \\ &= f_u(Z(u,y)) - f_u(px_0) \\ &= f_u(h_u^{-1}(T'_u(y))) - f_u(px_0) = T'_u(y). \end{aligned}$$

This proves our assertion.

If we now identify P with A/A_0, then the divisor $D(y)$ will generate an analytic system of divisors in $\mathfrak{G}_0(V)$, "generically" parametrized by P in the sense that P is the parameter variety of this system for all generic points of P over K'; furthermore, if $D(y_0)$ is any specialization of $D(y)$ over any specialization $y \rightarrow y_0$ over K', then we have $G(D(y_0)) = y_0$. However, this system gives a one-to-one representation of the group $\mathfrak{G}_0(V)/\mathfrak{G}_l(V)$ only "in general," since it is possible that for some special point y_0 in P there exist more than one specializations of $D(y)$ over the specialization $y \rightarrow y_0$ over K'.

Finally, going over once more the arguments in this section, it will be readily seen that we have in fact proved the following result: If H is a homomorphism of any Abelian variety B into $J(u)$ and if B is defined over K and H is defined over $K(u)$, then $H(B)$ is contained in $J'_1(u)$. Thus we can characterize the Picard variety P as the "largest Abelian subvariety" in $J(u)$ which is defined over K. As mentioned in section 1, this result is part of a purely algebraic theory of Picard varieties for algebraic varieties over an arbitrary ground field, which will be developed in a forthcoming paper.

THE JOHNS HOPKINS UNIVERSITY.

ON k-FOLD IRREDUCIBILITY OF MAPPINGS.*

By G. T. Whyburn.

1. **Introduction.** In an earlier paper [1] it was shown by the author that a mapping $f(A) = B$ on a compact set A is (strongly) irreducible if and only if the points x in A satisfying $x = f^{-1}f(x)$ are dense in A. In this paper a similar characterization of the property of being irreducibly k-fold will be established. As in the earlier theorem, the proof can be made to hinge on the upper semi-continuity of a strategically defined diameter function.

A mapping $f(A) = B$ is said to be *irreducibly k-fold*, k a positive integer, on a closed set X in A provided that for each $y \varepsilon B$, $X \cdot f^{-1}(y)$ contains at least k points but if X' is any closed proper subset of X, then $X' \cdot f^{-1}(y)$ contains less than k points for at least one $y \varepsilon B$. For $k = 1$ this is identical with ordinary irreducibility in the strong sense as referred to above. In what follows we shall assume that all spaces used are separable and metric. For any set Z, and positive number r, $V_r(Z)$ will denote the r-neighborhood of Z, i. e., the set of all points of the space at a distance $< r$ from the set Z.

2. **The function $e_k(x)$.** For any set A in a metric space and any positive integer k, let us define $e_k(A) = \text{g.l.b.}\,[\max \delta(A_i)]$ for all decompositions $A = \sum_1^k A_i$ of A into k non-empty subsets A_i. It will be noted at once that a set A is of power $\leqq k$ if and only if $e_k(A) = 0$. Also we remark that if k were allowed to assume the values ω (arbitrary finite number) and $\aleph_0$ we would have $e_\omega(A) = 0$ for all totally bounded sets A and, in a separable space, $e_{\aleph_0}(A) = 0$ for all sets A.

A mapping $f(X) = Y$ is said to generate an upper semi-continuous decomposition of X provided the decomposition of X into the sets $[f^{-1}(y)]$, $y \varepsilon Y$ is upper semi-continuous [2], i. e., the union of all sets of this collection contained in any open set in X is an open set in X.

Theorem. *If the mapping $f(X) = Y$ generates an upper semi-continuous decomposition of X, the function $e_k(x) \equiv e_k[f^{-1}f(x)]$, $x \varepsilon X$, is upper semi-continuous.*

* Received April 21, 1952.

Proof. Let N be any real number and let U be the set of all $x \varepsilon X$ for which $e_k(x) < N$. We have to show that N is open. To that end let $x \varepsilon U$. If $3d = N - e_k(x)$ there exists a decomposition

$$A = f^{-1}f(x) = \sum_1^k A_i,$$

with $\delta(A_i) < N - 2d$ for $i = 1, \cdots, k$. This gives $V_d(A) = \sum_1^k V_d(A_i)$ with $\delta[V_d(A_i)] < \delta(A_i) + 2d < N$, for $i = 1, \cdots, k$. Now by hypothesis the union V of all sets $f^{-1}(y)$, $y \varepsilon Y$, contained in $V_d(A)$ is open. Further, any such set admits the decomposition

$$f^{-1}(y) = f^{-1}(y) \cdot V_d(A) = \sum_1^k f^{-1}(y) \cdot V_d(A_i)$$

with $\delta[f^{-1}(y) \cdot V_d(A_i)] < N$, $i = 1, 2, \cdots, k$, so that $e_k(v) < N$ for all $v \varepsilon V$. Accordingly U is open because $x \varepsilon V \subset U$.

3. Characterization of k-fold irreducibility. A mapping $f(X) = Y$ is at least k-fold provided $f^{-1}(y)$ contains at least k points for each $y \varepsilon Y$.

THEOREM. *If X is compact, an at least k-fold mapping $f(X) = Y$ is irreducibly k-fold on X if and only if the set of all $x \varepsilon X$ such that $f^{-1}f(x)$ consists of exactly k-points is dense in X.*

Proof. If the latter condition is satisfied, then for any closed proper subset X' of X, $X - X'$ would contain a point x for which $f^{-1}f(x)$ consists of just k points and thus so that $f^{-1}(x)$ has at most $k - 1$ points in X'. Hence f is irreducibly k-fold on X.

On the other hand suppose f is irreducibly k-fold on X. Let ϵ be any positive number and let U be any open set in X of diameter $< \epsilon$. Then since f fails to be at least k-fold on $X - U$, there exists a $y \varepsilon Y$ such that $X - U$ contains at most $k - 1$ points of $f^{-1}(y)$ and thus such that $f^{-1}(y)$ is contained in U plus the union of $k - 1$ additional points of X. Thus $f^{-1}(y)$ admits a decomposition into k sets each of diameter $< \epsilon$ so that $e_k(x) < \epsilon$ for each $x \varepsilon f^{-1}(y)$. Thus any open set U contains a point for which $e_k(x) < \epsilon$. By the upper semi-continuity of $e_k(x)$ established in § 2 and the compactness of X, it follows that the set of zeros of $e_k(x)$ is dense in X; and it has already been noted that the zeros of $e_k(x)$ are precisely the points x for which $f^{-1}f(x)$ consists of exactly k points.

4. Existence theorem. In general, for a mapping the property of being at least k-fold on a set is not inducible for $k > 1$. Hence the standard procedure, based on the Brouwer Reduction Theorem, for showing the existence of a set on which a property is present irreducibly is not available for $k > 1$ although it yields the expected result in case $k = 1$. However for finite-to-one mappings this procedure does apply as will now be shown.

LEMMA. *For finite-to-one mappings the property of being at least k-fold on a set is inducible.*

For let $f(X) = Y$ be a finite-to-one mapping and let $A_1 \supset A_2 \supset A_3 \supset \cdots$ be a monotone decreasing sequence of non-empty compact sets in X on each of which f is at least k-fold and let A be their intersection. If for some $y \varepsilon Y$, $A \cdot f^{-1}(y)$ contained $< k$ points, then since $f^{-1}(y)$ contains only a finite number of points and any point of $f^{-1}(y)$ not in A would fail to be in A_n for some n, then for n sufficiently large $A_n \cdot f^{-1}(y)$ would contain less than k points. As this contradicts our supposition, the lemma is proven.

The Brouwer Reduction Theorem now yields at once the following

THEOREM. *If X is compact and $f(X) = Y$ is any at least k-fold finite-to-one mapping, there exists a closed subset A of X on which f is an irreducibly k-fold mapping onto Y.*

For mappings which are not finite to one the existence question requires further study and it is proposed to deal with this in a later paper.

UNIVERSITY OF VIRGINIA.

BIBLIOGRAPHY.

[1] G. T. Whyburn, "On irreducibility of transformations," *American Journal of Mathematics*, vol. 61 (1939), pp. 820-822.

[2] R. L. Moore, "Concerning upper semi-continuous collections of continua," *Transactions of the American Mathematical Society*, vol. 27 (1925), pp. 416-428. Also see Alexandroff-Hopf, *Topologie* I, Berlin, 1935 (Springer), and earlier references found therein.

REPRESENTATIONS OF PRIMES BY QUADRATIC FORMS.*

By N. C. Ankeny.[1]

Hecke proved that there exist infinitely many rational primes q such that

$$q = a^2 + b^2 \tag{1}$$

and $b^2 = o(q)$, as $q \to \infty$ (cf. [3]). In this paper we shall prove on the assumption of the Extended Riemann Hypothesis (E. R. H.) that there exist infinitely many rational primes q such that

$$q = a^2 + b^2 \tag{2}$$

and $b = O(\log q)$.

Hecke's method was the following: let α be any integer in the Gaussian Field F ($F = R(i)$, where R consists of the rational numbers). Now define the character $\chi(\alpha, n) = (\alpha/\bar{\alpha})^{2n}$, where n is any rational integer. Define

$$L(s, n) = \sum_{(\alpha)} \chi(\alpha, n)/N\alpha^s, \qquad \mathrm{Re}(s) > 1,$$

where (α) runs over all integral ideals in F. These new functions $L(s, n)$ define a new type of L-series which have many properties quite similar to the Dirichlet L-series. Hecke proved that the L-series had an Euler product and a functional equation.

Let $(P/\bar{P}) = e^{i\theta_P}$, where P runs over all prime numbers in F. Then, with the ordering of the prime ideals by the norms, $N_{FR}(P_i) \leqq {}_{FR}(P_{i+1})$

$$\theta_{P_1}, \theta_{P_2}, \theta_{P_3}, \cdots \tag{3}$$

are equidistributed (mod 2π). From this fact (1) immediately follows.

In the following we shall go further and prove a refinement of Hecke's results on the assumption of the E. R. H. We shall make use of the work of A. Selberg on the zeros of Dirichlet L-series (cf. [6], [7]).

We shall state here the two main theorems with regard to the equidistribution of (3).

Let $|\bar{u}|$ be defined as the absolute value of $\bar{u}$, where $u \equiv \bar{u} (\mathrm{mod}\, 2\pi)$,

* Received December 11, 1951; revised April 31, 1952.

[1] Prepared under the auspices of the Office of Naval Research.

$-\pi \leqq \bar{u} \leqq \pi$. Define $f_1 = f_1(\theta, T, \lambda) = 1$ if $|\lambda - \theta| \leqq (\log T)^{1+\epsilon} T^{-\frac{1}{2}}$; $f_1 = 0$ otherwise, where $\epsilon > 0$ and $T > T_0(\epsilon)$.

Let $N(P)$ be the norm from F to R of the ideal P.

THEOREM I. *If P runs over all prime ideals in F, then, for any λ,*

$$\sum_P f_1(\theta_P, T, \lambda)(\log N(P)) e^{-N(P)/T}$$
$$= 1/\pi T^{\frac{1}{2}}(\log T)^{1+\epsilon} + o(T^{\frac{1}{2}}(\log T)^{1+\epsilon}), \text{ as } T \to \infty.$$

Define $f_2 = f_2(\theta, T) = 1$ if $|\theta| \leqq c_1(\log T/T^{\frac{1}{2}})$; $f_2 = 0$ otherwise, where c_1 is a sufficiently large absolute constant. (This will be determined in Section **3**.)

THEOREM II. *If P runs over all prime ideals in F, then*

$$\sum_P f_2(\theta_P, T)(\log N(P)) e^{-N(P)/T} > T^{\frac{1}{2}} \log T, \qquad T \to \infty.$$

The result (2) is derived in Section **3** from Theorem II.

Theorem I and II can be generalized to any quadratic extension of the rationals and this will be done in Section **4**. The result (2) can also be generalized (see Theorem III).

I would like to thank A. Selberg for suggestions on the presentation of this manuscript as well as actual sharpening of certain of the results.

Sections **1** and **2** contain prelimnary lemmas. Section **3** deals with the proofs of Theorems I and II. Section **4** gives an outline of the generalization of these results to other quadratic extensions of the rationals.

1. Define $g(x, \delta_1, \delta_2) = g(x)$ as follows:

(a) $g(x + 2\pi) = g(x)$, (b) $g(-x) = g(x)$,
(c) $g(x) = 1, 1 + 1/\delta_1(\delta_2 - x)$

for $0 \leqq x \leqq \delta_2$, $\delta_2 \leqq x \leqq \delta_1 + \delta_2$, $\delta_2 + \delta_1 \leqq x \leqq \pi$, respectively, where $1/100 > \delta_2 \geqq \delta_1 > 0$. Then

$$g(x) = a_0 + \sum_{n=1}^{\infty} a_n \cos(nx), \tag{4}$$

where

$$2\pi a_0 = \delta_1 + 2\delta_2, \qquad \pi a_n = (\delta_1 n^2)^{-1}(\cos n(\delta_1 + \delta_2) - \cos n\delta_1). \tag{5}$$

We note that

$$|a_n| \leqq 5 \min(1/(\delta_1 n^2), \delta_2). \tag{6}$$

LEMMA 1. *If p runs over all rational primes $\equiv 1 \pmod 4$, T is a large positive integer, then*

$$\sum_p \Lambda(p) e^{-p/T} = \tfrac{1}{2}T + o(T)$$

as $T \to \infty$, where $\Lambda(p) = \log p$.

We shall not prove Lemma 1 as it is a standard theorem on primes in an arithmetic progression. Let us note here that, if p is an odd rational prime, $p \equiv 1 \pmod 4$ holds if and only if p is the norm of a prime ideal in F.

2. In the following let (α) run over all integral ideals in the Gaussian field F. Define $\chi(\alpha, n) = (\alpha/\bar{\alpha})^{2n}$, where n is any rational integer. Then $\chi(\alpha, n) = \chi(i\alpha, n) = \chi(-i\alpha, n) = \chi(-\alpha, n)$, so that $\chi(\alpha, n)$ depends only upon the ideal (α) and upon n. Further, $\chi(\alpha\beta, n) = \chi(\alpha, n)\chi(\beta, n)$. If $L(s, n) = \Sigma\, \chi(\alpha, n)(N_{FR}(\alpha))^{-s}$, $\mathrm{Re}(s) > 1$, then $L(s, n)$ satisfies the following conditions:

(A) $$L(s, n) = \prod_q (1 - \chi(q, n)(N_{FR}q)^{-s})^{-1}, \qquad \mathrm{Re}(s) > 1,$$

where q runs over all prime ideals in F.

(B) If $$R(s) = (\pi/2)^{-s/2} L(s, n) \Gamma(s + 2n),$$

then $R(s)$ is meromorphic in the entire S-plane (actually, regular in the entire S-plane if $n \neq 0$), and $R(s) = WR(1-s)$, where $|W| = 1$. (See Hecke [2], [3] and [4] for proofs. Also see Landau [5] for a proof.)

We assume $L(s, n) \neq 0$, for $\mathrm{Re}\, s > \frac{1}{2}$ and all n; we shall call this assumption the Extended Riemann Hypothesis.

Let p run over all rational primes which are the norms of prime ideals in F (i.e., $p = a^2 + b^2$ for some positive rational integers a and b). Let $\exp(i\theta_p) = ((a + bi)(a - bi)^{-1})^2$.

LEMMA 2. *If T is a large positive number and a a positive rational integer, then*

$$\sum_p (\cos n\theta_p) \Lambda(p) e^{-p/T} = O(T^{\frac{1}{2}} \log(n + 3)).$$

The proof of this lemma will not be given. The proof for Dirichlet L-series was given by Ankeny (cf. [1]), and the difference between these two types of L-series has no real relevance to the proof. The results of A. Selberg on the zeros on the zeros of Dirichlet L-series can be extended to this new type of L-series, and Selberg's results were the essential tools for the author's result in [1].

3. We are now in a position to prove Theorems I and II. First,

$$(7) \qquad g(\theta_p, \delta_1, \delta_2) = \sum_{n=0}^{\infty} a_n \cos n\theta_p,$$

where $g(x, \delta_1, \delta_2)$ was defined in Section 1, and θ_p in Section 2. Multiplying both sides of (7) by $\Lambda(p)e^{-p/T}$ and summing over p, we obtain

$$(8) \qquad \sum_p g(\theta_p, \delta_1, \delta_2)\Lambda(p)e^{-p/T} = a_0 \sum_p \Lambda(p)e^{-p/T} + \sum_{n=1}^{\infty} a_n \sum_p (\cos n\theta_p)\Lambda(p)e^{-p/T}$$

$$= a_0 \sum_p \Lambda(p)e^{-p/T} + O\Big(\sum_{n=1}^{\infty} | a_n | T^{\frac12} \log(n+3)\Big),$$

by Lemma 2. Now

$$(9) \qquad \sum_{n=1}^{\infty} | a_n | \log(n+3)$$

$$= \sum_{n \leqq (\delta_1\delta_2)^{-1/2}} | a_n | \log(n+3) + \sum_{n > (\delta_1\delta_2)^{-1/2}} | a_n | \log(n+3),$$

and by (6)

$$\leqq 5\delta_2 \sum_{n \leqq (\delta_1\delta_2)^{-1/2}} \log(n+3) + 5/\delta_1 \sum_{n > (\delta_1\delta_2)^{-1/2}} \log(n+3)/n^2$$

$$\leqq 20\delta_2^{\frac12}\delta_1^{-\frac12} \log 1/\delta_1\delta_2.$$

Hence, substituting (9) into (8),

$$(10) \qquad \sum_p g(\theta_p, \delta_1, \delta_2)\Lambda(p)e^{-p/T}$$

$$= \tfrac12\pi^{-1}(\delta_1 + 2\delta_2) \sum_p \Lambda(p)^{-p/T} + O(T^{\frac12}\delta_2^{\frac12}\delta_1^{-\frac12} \log \delta_1\delta_2).$$

The latter formula will yield the proof of both Theorems I, II and of (2) by certain selection of $\delta_,$, δ_2.

If

$$(11) \qquad \delta_1 = (\log T)^{1-\epsilon/2}/T^{\frac12}, \qquad \delta_2 = (\log T)^{1+\epsilon}/T^{\frac12}$$

where ϵ is some arbitrarily small number, then

$$\sum_p g(\theta_p, \delta_1, \delta_2)\Lambda(p)e^{-p/T} = \delta_2 T/\pi + \tfrac12\delta_1 T + o(\delta_2 T) + O(T^{\frac12}(\log T)^{1+3/4\epsilon})$$

$$= 1/\pi T^{\frac12}(\log T)^{1+\epsilon} + O(T^{\frac12}(\log T)^{1+3/4\epsilon}) + o(T^{\frac12}(\log T)^{1+\epsilon}).$$

Let $f_1(\theta, T)$ be 1 if $| \theta | \leqq \delta_2$ and 0 if $\delta_2 \leqq x \leqq 2\pi - \delta_2$. Then

$$g(x, \delta_1, \delta_2 - \delta_1) \leqq f(x, \delta_2) \leqq g(x, \delta_1, \delta_2)$$

and, using the values (11) of δ_1, δ_2,

$$\sum_p f_1(\theta_p, (\log T)^{1+\epsilon}/(T^{\frac12}))\Lambda(p)e^{-p/T} = T^{\frac12}/(\pi)(\log T)^{1+\epsilon} + o(T^{\frac12}(\log T)^{1+\epsilon}).$$

Similarly, we could extend the result of $f_1(\theta, (\log T)^{1+\epsilon}/(T^{\frac{1}{2}}), \lambda)$ defined in the introduction. This results by merely changing the definition of $f_2(\theta, \delta_1, \delta_2)$ by a translation on θ of length λ. We shall leave this modification to the reader.

We have now completed the proof of Theorem I. To prove Theorem II and (2), we proceed as follows: Let $f_2(\theta, T)$ be 1 if $|\theta| < c_1 T^{-\frac{1}{2}} \log T$, and f_2 be 0 otherwise, where c_1 is a constant that will be chosen later. Then, in (10), letting $\delta_1 = \delta_2 = c_2 T^{-\frac{1}{2}} \log T$ and using Lemma 1

$$\sum_p g(\theta_p, c_2 T^{-\frac{1}{2}} \log T, c_2 T^{-\frac{1}{2}} \log T) \Lambda(p) e^{-p/T} = 3\pi^{-1} c_2 T^{\frac{1}{2}} \log T + O(T^{\frac{1}{2}} \log T),$$

$> T^{\frac{1}{2}} \log T$, if c_2 is chosen sufficiently large. If $c_1 = 3c_2$, then

$$f_2(x, T) \geqq g(x, c_2 T^{-\frac{1}{2}} \log T, c_2 T^{-\frac{1}{2}} \log T),$$

so that

$$\sum_p f_2(\theta_p, T) \Lambda(p) e^{-p/T} \geqq \sum_p g(\theta_p, c_2 T^{-\frac{1}{2}} \log T, c_2 T^{-\frac{1}{2}} \log T) > T^{\frac{1}{2}} \log T. \tag{12}$$

From (12) we can deduce (2) provided we can show that

$$\sum_{p > c_4 T} f_2(\theta_p, T)(p) e^{-p/T} < \tfrac{1}{2} T^{\frac{1}{2}} \log T$$

for c_4 sufficiently large. To this end, define

$$H(U) = \sum_{p \leqq U} f_2(\theta_p, T) \Lambda(p). \tag{13}$$

We note that $f_2(x, U) \leqq g(x, c_1 T^{-\frac{1}{2}} \log T, c_1 T^{-\frac{1}{2}} \log T)$ for $U \geqq T$. Hence, if $\delta_1 = \delta_2 = c_1 T^{-\frac{1}{2}} \log T$ in (10), then

$$\begin{aligned} e^{-1} H(U) = e^{-1} \sum_{p \leqq U} f_2(\theta_p, U) \Lambda(p) &\leqq \sum_p f_2(\theta_p, U) \Lambda(p) e^{-p/U} \\ &\leqq \sum_p g(\theta_p, c_1 T^{-\frac{1}{2}} \log T, c_1 T^{-\frac{1}{2}} \log T) \Lambda(p) e^{-p/U} \\ &= (3c_1/2\pi) T^{-\frac{1}{2}} \log T \sum_p \Lambda(p) e^{-p/U} + O(U^{\frac{1}{2}} \log T) \\ &= (3c_1/2\pi) U T^{-\frac{1}{2}} \log T + O(U^{\frac{1}{2}} \log T) < c_5 U T^{-\frac{1}{2}} \log T. \end{aligned} \tag{14}$$

Hence, by partial summation,

$$\sum_{p \geqq c_4 T} f_2(\theta_p, T) \Lambda(p) e^{-p/T} = \sum_{n \geqq c_4 T} H(n)(e^{-n/T} - e^{-(n+1)/T}), \tag{15}$$

which, by (14), is

$$\begin{aligned} = T^{-1} \int_{c_4 T}^{\infty} H(U) e^{-U/T} dU &\leqq T^{-1} \int_{c_4 T}^{\infty} (c_5 T^{-\frac{1}{2}} \log T) U e^{-U/T} dU \\ &= c_5 (c_4 + 1) e^{-c_4} T^{\frac{1}{2}} \log T \end{aligned}$$

$< \frac{1}{2}T^{\frac{1}{2}} \log T$ if c_4 is sufficiently large. Hence, by (12),

$$(16) \quad \sum_{p < c_4 T} f(\theta_p, T)(p) e^{-p/T} > T \log T - \sum_{p \geqq c_4 T} f_2(\theta_p) \Lambda(p) e^{-p/T} > \tfrac{1}{2} T \log T.$$

Hence there exists a prime q such that $q < c_4 T$ and

$$(17) \qquad |\theta_q| < c_1 T^{-\frac{1}{2}} \log T.$$

As $q = a^2 + b^2$, $a > b > 0$,

$$\exp \pm i\theta_q = ((a + bi)(a - bi)^{-1}), \exp i\theta_q = \cos \theta_q + i \sin \theta_q.$$

Hence, by (17), for T sufficiently large,

$$bq^{-\frac{1}{2}} = \sin \theta_q/2 \leqq c_6 T^{-\frac{1}{2}} \log T \leqq c_7 q^{-\frac{1}{2}} \log q,$$

and so $b = O(\log q)$. This implies that there are infinitely many primes satisfying (2); in fact, from (12) we can obtain a lower bound for their density.

4. One can proceed to generalize Theorems I and II to all quadratic imaginary extensions of the rationals. If the class number h of F is greater than one, we must define our characters for ideals that are not necessarily principal. In F the group of classes is a finite Abelian group. Therefore, the group has a finite number of generators, and so every ideal in F can uniquely be put in the form $= (\alpha) A_1^{i_1} A_2^{i_2} \cdots A_r^{i_r}$, where $0 \leqq i < h_1, \cdots,$ $0 \leqq i_r < h_r$, with $h = h_1 h_2 \cdots h_r$, and $A_1, \cdots, A_r$, are certain ideals in F. Also $A_i^{h_i} = (\beta_i)$ with $\beta_i \in F$.

Let $\beta_i^* = \beta_i^{1/h_i}$; this definition of β_i^* is unique up to an h_i-th root of unity. Define

$$\chi_j(A, n) = (\alpha/\bar{\alpha} \prod_{j=1}^{r} (\beta_j^*/\bar{\beta}_j^*))^{en}$$

where e is half the number of roots of unity in F. Although $\beta_i^*/\bar{\beta}_i^*$ is not uniquely determined, whatever choice of β_i^* we make, determines χ_j. As $h_1 h_2 \cdots h_r = h$, we have at most h different characters.

To establish our theorems we must also introduce the characters $\tau_0, \tau_1, \cdots,$ τ_{h-1} which form the character group of the ideal classes in F.

If we define $\exp(i\theta_{pj}) = \chi_j(P, 1)$, then we can proceed as before to show that the θ_{pj} are strongly equidistributed in the sense of Theorem I. Furthermore, using the τ's, we can show that the θ_{pj} are strongly equidistributed, where p runs over all prime ideals in a given class in F. We will state the interpretation of (2) in this case.

Let $H(x, y) = \alpha x^2 + \beta xy + \alpha y^2$ be a positive definite binary quadratic

form associated with an ideal class in F. (See Landau [6] for the relationship between binary quadratic forms and ideal classes in quadratic fields.)

THEOREM III. *There are infinitely many primes q such that*

$$q = H(a, b), \qquad b = O(\log q).$$

If F is a quadratic real extension of R, then we define our characters as follows: If $\alpha \varepsilon F$, $\alpha \neq 0$, let $\chi(\alpha, n) = \exp(\pi i\, n (\log \alpha)(\log \alpha)^{-1} (\log \epsilon)^{-1})$, where α is the conjugate of α, and ϵ is the fundamental unit in F.

The generalization for defining a character for an element to defining a character for an ideal is as follows: Let $A = (\alpha) A_1^{i_1} A_2^{i_2} \cdots A_r^{i_r}$, $A_j^{h_j} = (\beta_j)$. Define $\beta_j^* = (\beta_j)^{1/h_j}$, $\bar{\beta}_j^* = (\bar{\beta}_j)^{1/h_j}$, where both β_j^* and $\bar{\beta}_j^*$ are real. Then

$$\chi_{l_j}(A_j, n) = \exp \pi i\, n (\log \beta_j^* / \log \beta_j (\log \epsilon)^{-1} + 2 l_j / h_j),$$

where l_j runs through a set of residues $(\operatorname{mod} h_j/(2, h_j))$.

The condition in Theorem III that $f(x, y)$ need be associated with an ideal class in a quadratic imaginary extension of the rationals can now be reduced to an association with an ideal class in any quadratic extension of the rationals.

THE INSTITUTE FOR ADVANCED STUDY.

BIBLIOGRAPHY.

[1] N. C. Ankeny, "The least quadratic non-residue," *Annals of Mathematics*, vol. 55 (1952), pp. 65-72.

[2] E. Hecke, "Über die L-Funktionen und den Dirichletschen Primzahlsatz für einen beliebigen Zahlkörper," *Göttinger Nachrichten* (1917), pp. 299-318.

[3] ———, "Eine neue Art von Zetafunktionen und ihre Beziehungen zur Verteilung der Primzahlen," *Mathematische Zeitschrift*, vol. 1 (1918), pp. 357-376.

[4] ———, "Eine neue Art von Zetafunktionen und ihre Beziehungen zur Verteilung der Primzahlen," *Mathematische Zeitschrift*, vol. 6 (1920), pp. 11-51.

[5] E. Landau, "Verallgemeinerung eines Pólyaschen Satzes auf algebraische Zahlkörper," *Göttinger Nachrichten* (1918), pp. 478-488.

[6] A. Selberg, "On the zeros of Riemann's zeta-function," *Norske Videnskapsakademi Skrifter*, no. 10 (1942).

[7] ———, "Contributions to the theory of Dirichlet's L-functions," *ibid.*, no. 3 (1946).

ON THE L^2-SPACE OF A LIE GROUP.*

By G. D. Mostow.[1]

1. Introduction. The purpose of this paper is to prove that the L^2-space of a Lie group G has a base of functions which are analytic throughout G. For a compact group this follows readily from the Peter-Weyl Theorem together with the theorem that a closed subgroup of an analytic group is analytic. In order to establish the theorem for a general Lie group G, it is proved that any Lie group is analytically the direct product of a compact subgroup K and a euclidean subspace E; that, moveover, left Haar measure on G is the product measure of Haar measure on K by ordinary euclidean measure on E (relative to suitable coordinates). The main result follows from this. It is also proved that any Lie group can be embedded analytically with non-vanishing Jacobian in euclidean space.

2. Definitions and Preliminaries. We employ the terms "analytic submanifold," "analytic isomorphism of an analytic manifold," "analytic subgroup" in the sense defined in [1] (cf. pp. 74, 85, 100, 107).

Definition. Let ϕ be an analytic mapping of A onto B. Let A_0, B_0 be analytic submanifolds of A, B respectively. A_0 and B_0 are called ϕ-related if and only if A_0 is a connected component of $\phi^{-1}(B_0)$.

Definition. Let G be an analytic group, let $A_1, \cdots, A_n$ be analytic submanifolds of G, and let θ denote the mapping $(a_1, a_2, \cdots, a_n) \to a_1 \cdot a_2 \cdots\cdot a_n$ of $A_1 \times A_2 \times x \cdots \times A_n$ into G. We say

a) $G = A_1 \cdot A_2 \cdot \cdots \cdot A_n$ if θ is onto;

b) $G = A_1 \cdot A_2 \cdot \cdots \cdot A_n(\mathrm{UN})$ if θ is one-to-one onto;

c) $G = A_1 \cdot A_2 \cdot \cdots \cdot A_n(\mathrm{TP})$ if θ is a homeomorphism onto;

d) $G = A_1 \cdot A_2 \cdot \cdots \cdot A_n(\mathrm{AN})$ if θ is an analytic isomorphism onto.

Lemma 1. *Let G, G^* be analytic groups, and assume*

1. $G^* = A_1^* \cdot A_2^* \cdot \cdots \cdot A_n^*(\mathrm{UN})$;
2. $G = A_1 \cdot A_2 \cdot \cdots \cdot A_n(\mathrm{AN})$;

* Received November 13, 1950; revised April 16, 1952.

[1] Prepared under contract NR 045-004 with the Office of Naval Research.

3. *Each* A_i *and* A_i^* *are* ϕ*-related, where* ϕ *is a locally* 1-1 *homomorphism from* G *onto* G^* (or vice versa).

Then $G^* = A_1^* \cdot A_2^* \cdot \cdots \cdot A_n^*(\mathrm{AN})$.

Proof. This lemma will be proved only for the case that ϕ is a homomorphism of G onto G^*. In the reverse case, the proof is essentially the same. We show that there is an open set about any element $x^* \varepsilon G^*$ on which the mapping

$$\theta^*:\ a_1^* a_2^* \cdots a_n^* \to (a_1^*, a_2^*, \cdots, a_n^*),\quad a_i^* \varepsilon A_i^*$$

is an analytic isomorphism. Let $x^* = b_1^* \cdot b_1^* \cdots b_n^*$, $b_i^* \varepsilon A_i^*$ $(i = 1, \cdots, n)$ and let $x \varepsilon G$ be ϕ-related to x^*. Then $x = b_1 \cdot b_2 \cdot \cdots \cdot b_n$, where $b_i \varepsilon A_i$ and b_i is ϕ-related to b_i^* $(i = 1, \cdots, n)$. Let V_i be an open neighborhood of b_i^* in A_i^* on which ϕ is one-to-one and let $U = V_1 \cdot V_2 \cdot \cdots \cdot V_n$. Let Φ denote the mapping $(a_1, a_2, \cdots, a_n) \to (\phi(a_1), \cdots, \phi(a_n))$ of $V_1 \times V_2 \times \cdots \times V_n$ onto $\phi(V_1) \times \cdots \times \phi(V_n)$. Since ϕ is an analytic isomorphism on U, Φ is an analytic isomorphism. Let ϕ^* denote the mapping of the open set $\phi(U)$ onto U which is the inverse of ϕ. Inasmuch as $\theta^* = \Phi\theta\phi^*$ (ϕ^* maps first), where θ is the mapping $a_1 a_2 \cdots a_n \to (a_1, a_2, \cdots, a_n)$ of G onto $A_1 \times A_2 \times \cdots \times A_n$, and inasmuch as each of ϕ^*, θ, Φ are analytic isomorphisms, θ^* is an analytic isomorphism on the open set $\phi(U)$. Since x^* was an arbitrary point of G^*, θ^* is an analytic isomorphism and $G^* = A_1^* \cdot A_2^* \cdot \cdots \cdot A_n^*(\mathrm{AN})$.

LEMMA 2. *Let* G *be an analytic group, and let* A, B *be analytic subgroups. If* $G = A \cdot B(\mathrm{UN})$, *then* $G = A \cdot B(\mathrm{AN})$.

Proof. Let $\mathfrak{G}$, $\mathfrak{A}$, $\mathfrak{B}$ denote the Lie algebras of G, A, B respectively, and let $X_1, \cdots, X_a$ and $Y_1, \cdots, Y_b$ be bases for $\mathfrak{A}$ and $\mathfrak{B}$ respectively. Then $X_1, X_2, \cdots, X_a$, $Y_1, \cdots, Y_b$ form a base for $\mathfrak{G}$. Let

$$U = \{z \mid z = \exp s^1 X_1 \cdots \exp s^a X_a,\ |s^i| < \epsilon,\ i = 1, \cdots, a\},$$

$$V = \{z \mid z = \exp t^1 Y_1 \cdots \exp t^b Y_b,\ |t^i| < \epsilon,\ i = 1, \cdots, b\}.$$

Then for ϵ suitably small, U, V, and UV are open neighborhoods of the identity in A, B, and G respectively and the mapping $\phi: (u, v) \to u \cdot v$ is an analytic isomorphism of $U \times V$ onto $U \cdot V$ (cf. [5], p. 193). Select ϵ so that these conditions hold.

Let L_x and R_x for $x \varepsilon G$ denote respectively left and right group translation by x. Let L_a^* for $a \varepsilon A$ and R_b^* for $b \varepsilon B$ denote the transformations $(x, y) \to (ax, y)$ and $(x, y) \to (x, yb)$ of $A \times B$ onto itself. Since A, B,

and G are analytic groups, L_x, R_x, L_a^*, and R_b^* are analytic isomorphisms. Let (x, y) be any element of $A \times B$ and let θ denote the mapping $(a, b) \to ab$ of $A \times B$ onto G. On the open set (xU, Vy) we have $\theta = L_x R_y \phi L_x^{*-1} R_y^{*-1}$, so that θ is an analytic isomorphism. Thus θ is a regular analytic mapping at all points of $A \times B$ and hence is an analytic isomorphism.

As a consequence of Lemma 2, it is readily seen that the mapping $ab \to a_0 a \cdot b_0 b$ of G onto G is an analytic isomorphism.

3. Decomposition of Haar measure.

Definition. A measure on an analytic n-dimensional manifold is an alternating analytic n-form (cf. [1], p. 146).

Definition. A left Haar measure on an analytic group is a non-zero left invariant measure (cf. [1], p. 167).

Suppose that $G = A \cdot B(\mathrm{AN})$. In the sequel we let α and β denote the mappings $a \cdot b \to a$ of G onto A and $a \cdot b \to b$ of G onto B respectively. Suppose that w_A and w_B are measures on A and B respectively.

Definition. $w_A \times w_B = \delta\beta w_A \square \delta\beta w_B$ (cf. [1] p. 151 for meaning of notation). This is called "the product measure."

As is known (cf. [1], p. 166),

$$\int_{A\times B} h w_A \times w_B = \int_B \left(\int_A h w_A\right) w_B$$

for any function h on $A \times B$. If $G = A_1 \cdot A_2 \cdot \cdots \cdot A_n(\mathrm{AN})$ and $w_1, \cdots, w_n$ are measures on $A_1, \cdots, A_n$, we define $w_1 \times w_2 \times \cdots \times w_n$ analogously.

Let $L(g)$ and $R(g)$ denote left and right group translation by g respectively. If $G = A \cdot B(\mathrm{AN})$ where A and B are subgroups, let $L^*(a_0 b_0)$ denote the analytic isomorphism $ab \to a_0 a \cdot b_0 b$. Let $L'(a)$ denote restriction of $L(a)$ to A if $a \in A$, and let $L'(b)$ denote the restriction of $L(b)$ to B if $b \in B$.

LEMMA 3. *Suppose that A and B are subgroups of the analytic group G such that $G = A \cdot B(\mathrm{AN})$. Let ϕ and ψ denote left Haar measures on A and B respectively. Then $(\phi \times \psi)_{ab} = \delta L^*(a^{-1}b^{-1})_{ab}(\delta\alpha_e\phi_e \square \delta\beta_e\psi_e)$, where e is the identity element of G.*

Proof. By definition, $(\phi \times \psi)_{ab} = \delta\alpha_{ab}\phi_a \square \delta\beta_{ab}\psi_b$. Since ϕ and ψ are left Haar measures, $\phi_a = \delta L'(a^{-1})_a \phi_e$ and $\psi_b = \delta L'(b^{-1})_b \psi_e$. Since

$$L'(a^{-1})\alpha = \alpha L^*(a^{-1}b^{-1}), \qquad \delta L'(a^{-1}) = \delta L^*(a^{-1}b^{-1})\delta\alpha.$$

Consequently

$$\delta\alpha_{ab}\phi_a = \delta\alpha_{ab}\delta L'(a^{-1})_a\phi_e = \delta L^*(a^{-1}b^{-1})_{ab}\delta\alpha_e\phi_e.$$

Similarly $\delta\beta_{ab}\psi_b = \delta L^*(a^{-1}b^{-1})_{ab}\delta\beta_e\psi_e$. It follows that

$$(\phi \times \psi)_{ab} = \delta L^*(a^{-1}b^{-1})_{ab}\delta\alpha_e\phi_e \,\Box\, \delta L^*(a^{-1}b^{-1})_{ab}\delta\beta_e\psi_e$$
$$= \delta L^*(a^{-1}b^{-1})_{ab}(\delta\alpha_e\phi_e \,\Box\, \delta\beta_e\psi_e).$$

Let $\mathfrak{G}$ denote the Lie algebra of G, and let $\mathfrak{A}$, $\mathfrak{B}$ denote the Lie subalgebras of A and B respectively.

LEMMA 4. *Let w denote the left Haar measure determined by*

$$\delta\alpha_e\phi_e \,\Box\, \delta\beta_e\psi_e, \text{ i. e., } w_{ab} = \delta L(b^{-1}a^{-1})_{ab}(\delta\alpha_e\phi_e \,\Box\, \delta\beta_e\psi_e)$$

Let $T(b)$ denote the linear transformation induced on the factor space $\mathfrak{G}/\mathfrak{B}$ by the differential at e of the inner automorphism $L(b^{-1})R(b)$, where $b \in B$. Then $w = f \cdot \phi \times \psi$, where f is the function on G with

$$f(a \cdot b) = \det T(b) \equiv |\, T(b)\,|.$$

Proof. It follows directly from Lemma 3 that

$$\begin{aligned} w_{ab} &= \delta L(b^{-1}a^{-1})_{ab}\delta L^*(ab)_e(\phi \times \psi)_{ab} \\ &= |\, \delta L(b^{-1}a^{-1})_{ab}\delta L^*(ab)_e \,|\, (\phi \times \psi)_{ab} \\ &= |\, dL^*(ab)_e dL(b^{-1}a^{-1})_{ab} \,|\, (\phi \times \psi)_{ab} \\ &= |\, dL(b^{-1}a^{-1})_{ab} dL^*(ab)_e \,|\, (\phi \times \psi)_{ab}. \end{aligned}$$

To evaluate the determinant of the linear transformation $dL(b^{-1}a^{-1})\,dL^*(ab)$ on $\mathfrak{G}$, we select as a base for $\mathfrak{G}$ the union of a base for $\mathfrak{A}$ and a base for $\mathfrak{B}$.

On A, $L(b^{-1}a^{-1}) \cdot L^*(ab) = L(b^{-1})R(b)$, and on B, $L(b^{-1}a^{-1}) \cdot L^*(ab)$ is the identity transformation. Consequently

$$f(a, b) \equiv |\, dL(b^{-1}a^{-1})_{ab}dL^*(ab)_e \,| = \begin{vmatrix} T_m(b) & J \\ 0 & I \end{vmatrix} = |\, T_m(b)\,|,$$

where $T_m(b)$ is the matrix form of $T(b)$ relative to suitable choice of base in $\mathfrak{G}/\mathfrak{B}$, and I is an identity matrix. It follows that $f(a, b) = |\, T(b)\,|$.

COROLLARY 4 · 1. *If B is normal, then $f(a, b) = 1$.*

Proof. If $X \in \mathfrak{A}$ and $Y \in \mathfrak{B}$, then

$$dL(\exp - Y)\,dR(\exp Y)X = X[I + adY + (adX)^2/2! + \cdots] = X + Z,$$

where $Z \in \mathfrak{B}$. Since any element of B is a finite product of exponentials,

$dL(b^{-1})dR(b)(X) = X + Z$, where Z is some element of $\mathfrak{B}$ for all $b \in B$. It follows that $T(b)$ is the identity transformation for all $b \in B$, and so $f(a, b) = |T(b)| = 1$.

COROLLARY 4 · 2. *If B is compact or semi-simple, then $f(a, b) = 1$.*

Proof. $|T(b)|$ is a real one-dimensional representation of the compact (or semi-simple) connected group B. It follows that $|T(b)| = 1$ for all $b \in B$.

COROLLARY 4 · 3. *If the analytic subgroup determined by the Lie algebra $\mathfrak{B}$ in some analytic group with Lie algebra $\mathfrak{G}$ is compact, then $|T(b)| = 1$ for all $b \in B$.*

Proof. Let G' be an analytic group with Lie algebra $\mathfrak{G}$, let B' be the analytic subgroup determined by $\mathfrak{B}$, and assume B' is compact. Let π denote the local isomorphism between G and G'. Then $T(b) = T(\pi(b))$ when $b \in B$. Consequently, $|T(b)| = 1$ for all b in some neighborhood U of the identity in B. Since T is a homomorphism of B, $|T(b_1 b_2 \cdots b_n)| = 1$ when $b_1 \cdots b_n \in U$. Since B is connected, $\Sigma U^n = B$. Hence $|T(b)| = 1$ for all $b \in B$.

COROLLARY 4 · 4. *If* ad Y *is nilpotent for all $Y \in \mathfrak{B}$, then $f(a, b) = 1$.*

Proof. If $Y \in B$, then $T(\exp Y)$ is the exponential of a nilpotent transformation and therefore has determinant one. It follows that $|T(b)| = 1$ for all $b \in B$.

LEMMA 5. *Suppose that G and G^* are analytic groups and that π is a locally one-to-one homomorphism of G onto G^*. Assume that $G = M \cdot N$*(AN), *$G^* = M^* \cdot N^*$*(AN), *and that M, M^* and N, N^* are π-related. Suppose that w_M, w_N, w_G, w_{M^*}, w_{N^*}, w_{G^*} are measures on M, N, G, M^*, N^*, G^* respectively such that*

$$\delta\pi\delta\mu^* w_{M^*} = \delta\mu w_M, \qquad \delta\pi\delta\nu^* w_{M^*} = \delta\nu w_N, \qquad \delta\pi w_{G^*} = w_G,$$

where μ, ν and μ^*, ν^* denote the projections of G onto M, N and of G^* onto M^*, N^* respectively. Then $w_{G^*} = w_{M^*} \times w_{N^*}$ if and only if $w_G = w_M \times w_N$.

The proof follows directly from the fact that $d\pi$ is a regular onto mapping at each point and hence that $\delta\pi$ is an isomorphism of the ring of differential forms. Suppose G is an analytic group with measure w and $A_1, \cdots, A_n$ are submanifolds with measures $w_1, \cdots, w_n$.

Definition. $G = A_1 \cdot A_2 \cdots A_n$ $(w, w_1, \cdots, w_n)$ means

$$G = A_1 \cdot A_2 \cdots A_n(\text{AN}) \text{ and } w = w_1 \times w_2 \cdots \times w_n.$$

If $A_1, A_2, \cdots, A_n$ are analytic subgroups, then $G = A_1 \cdot A_2 \cdots A_n(\mathrm{m})$ means $G = A_1 A_2 \cdots A_n(\mathrm{AN})$ and left Haar measure on G is the product of left Haar measures on $A_1, A_2, \cdots A_n$.

If G and G^* are analytic groups, if w^* is a left Haar measure on G^*, and if π is a locally one-to-one homomorphism of G onto G^*, then $\delta\pi w^*$ is a left Haar measure on G. For $L^{-1}(g^*)\pi = \pi L^{-1}(g)$, where $g^* = \pi(g)$, implies

$$\delta\pi_g w^*_{g^*} = \delta\pi_g \delta L^{-1}{}_{g^*}(g^*) w^*_{e^*} = \delta L^{-1}{}_g(g)(\delta\pi_e w^*_{e^*}),$$

where e, e^* are the identity elements in G, G^* respectively. Thus $\delta\pi w^*$ is the left Haar measure determined by $\delta\pi_e w^*_{g^*}$. It follows that if w is any Haar measure on G, then w and $\delta\pi w^*$ differ only by a constant factor.

THEOREM 1. *Let G be an analytic group. Then $G = E \cdot K(w_G, w_E, w_K)$, where K is a compact subgroup, E is analytically isomorphic to a euclidean space, w_G and w_K are left Haar measures, and w_E is euclidean measure on E relative to suitable coordinates.*

Proof. We shall make use of topological decompositions due to Malcev [2] and Iwassawa [3].

Let $\tilde{G}$ denote the simply connected group of G, let π denote the natural homomorphism of $\tilde{G}$ onto G, and let $\tilde{L} \cdot \tilde{R}$ be a Levi decomposition of $\tilde{G}$ into the semi-direct product of a semi-simple subgroup $\tilde{L}$ and the radical $\tilde{R}$. Let $L = \pi(\tilde{L})$, let $L^* = \mathrm{Ad}(L)$, where Ad denotes the adjoint representation. Throughout our discussion, analytic subgroups of L^* and L that are denoted by X^* and X are Ad-related; analytic subgroups of $\tilde{G}$ and G that are denoted by $\tilde{X}$ and X are π-related.

By a theorem of Iwassawa [2] which is related to the representation of a matrix as the product of a triangular by an orthogonal matrix, $L^* = S^* \cdot C^*(\mathrm{TP})$ where S^* is a simply connected, solvable analytic subgroup and C^* is a compact analytic subgroup (cf. also [4]). From this one obtains readily that $L = S \cdot C(\mathrm{TP})$, $C = A \cdot K_1(\mathrm{TP})$ where K_1 is a compact subgroup and A is a vector subgroup central in C, $L = S \cdot C(\mathrm{TP})$, and the center of L is contained in C. (This last result follows from the fact that L/C is a covering space of the simply connected space $S^* = L^*/C^*$ and is therefore univalent.) Thus $\tilde{G} = \tilde{L} \cdot \tilde{R} = \tilde{S} \cdot \tilde{C} \cdot \tilde{R} = \tilde{S}(\tilde{C}\tilde{R})(\mathrm{TP})$, where $\tilde{C}\tilde{R}$ contains the center of $\tilde{G}$. It follows directly that $G = S \cdot (CR)(\mathrm{TP})$.

Now, by a theorem of Malcev, $CR = U_1 \cdot U_2 \cdot \cdots \cdot U_n \cdot (K_3 K_2 K_1)(\mathrm{TP})$, where 1) U_i is a subgroup isomorphic to the reals; 2) $U_i \cdots U_n \cdot K_3$ is a subgroup $(i = 1, \cdots, n)$ that is normal in $U_{i-1} \cdots U_n \cdot K_3$ $(i = 2, \cdots, n)$,

3) K_3, K_2 are compact subgroups of the solvable group AR which commute with each other and with K_1 (K_3 is in the closure of the commutator subgroup of AR); 4) $\tilde{K}_3\tilde{K}_2\tilde{K}_1 = \tilde{K}_3 \cdot \tilde{K}_2 \cdot K_1(\mathrm{TP})$. Thus

$$\begin{aligned}
\tilde{C}\tilde{R} &= (\tilde{U}_1\tilde{U}_2 \cdots \tilde{U}_n\tilde{K}_3) \cdot (\tilde{K}_2\tilde{K}_1)(\mathrm{TP}) \\
&= (\tilde{U}_1\tilde{U}_2 \cdots \tilde{U}_n\tilde{K}_3) \cdot (\tilde{K}_2\tilde{K}_1)(\mathrm{AN}) && \text{(Lemma 2)} \\
&= (\tilde{U}_1\tilde{U}_2 \cdots \tilde{U}_n\tilde{K}_3) \cdot (\tilde{K}_2\tilde{K}_1)(\mathrm{m}) && \text{(Corollary } 4 \cdot 3) \\
&= [\tilde{U}_1 \cdot (\tilde{U}_2 \cdots \tilde{U}_n\tilde{K}_3)] \cdot (\tilde{K}_2\tilde{K}_1)(\mathrm{m}) && \text{(Corollary } 4 \cdot 1) \\
&= [\tilde{U}_1 \cdot \tilde{U}_2 \cdots \tilde{U}_n \cdot \tilde{K}_3](\tilde{K}_2\tilde{K}_1)(\mathrm{m}) && \text{(Corollary } 4 \cdot 1) \\
&= \tilde{U}_1 \cdot \tilde{U}_2 \cdots \tilde{U}_n \cdot (\tilde{K}_3 \cdot \tilde{K}_2\tilde{K}_1)(\mathrm{m})
\end{aligned}$$

since the formation of the direct product of measure spaces is associative. But $\tilde{K}_3 \cdot \tilde{K}_2\tilde{K}_1 = \tilde{K}_3\tilde{K}_2\tilde{K}_1(\mathrm{m})$ by Corollary $4 \cdot 3$. Hence

$$\tilde{C}\tilde{R} = \tilde{U}_1 \cdot \tilde{U}_2 \cdots \tilde{U}_n \cdot (\tilde{K}_3\tilde{K}_2\tilde{K}_1)(\mathrm{m})$$

$$\text{and } CR = U_1, U_2, \cdots, U_n \cdot (K_3K_2K_1)(\mathrm{m})$$

by Lemmas 1 and 5. But $\tilde{G} = \tilde{L} \cdot \tilde{R}(\mathrm{m}) = (\tilde{S} \cdot \tilde{C}) \cdot \tilde{R}(\mathrm{m})$, by Corollary $4 \cdot 1$ and Corollary $4 \cdot 3$, hence $\tilde{G} = \tilde{S} \cdot (\tilde{C} \cdot \tilde{R})(\mathrm{m})$. Consequently, $G = S \cdot (CR)(\mathrm{m})$ by Lemmas 1 and 5.

As is well known for simply connected solvable groups, $S = V_1 \cdot V_2 \cdots V_m$, where each V_i is isomorphic to the reals and $V_{i+1} \cdots V_m$ is normal in $V_i \cdots V_n$ $(i = 1, \cdots, n)$. Hence, by repeated use of Corollary $4 \cdot 1$, we obtain

$$G = V_1 \cdot V_2 \cdots V_m \cdot U_1 \cdots U_n \cdot (K_3K_2K_1)(\mathrm{m}).$$

Let

$E = V_1V_2 \cdots V_mU_1 \cdots U_n$, $K = K_3K_2K_1$, and let $w_E = w_1 \times \cdots \times w_{m+n}$,

where w_i is left Haar measure in the i-th factor. Let $\mathfrak{R}_i$ denote the Lie subalgebra of the i-th factor $(i = 1, \cdots, m+n)$, let $\mathfrak{R}^{m+n}$ denote the Cartesian product $\mathfrak{R}_1 \times \cdots \times \mathfrak{R}_{m+n}$, and let θ denote the analytic isomorphism

$$(a_1, \cdots, a_{m+n}) \to \exp a_1, \cdot \exp a_2 \cdots \exp a_n, \qquad a_i \in \mathfrak{R}_i, \quad (i = 1, \cdots, m+n),$$

of $\mathfrak{R}^{m+n}$ onto E and let π_i denote the projection of $\mathfrak{R}^{m+n}$ onto $\mathfrak{R}_i$. Then

$$\delta\theta(w_1 \times w_2 \times \cdots \times w_{m+n}) = \delta\pi_1(u_1) \times \cdots \times \delta\pi_{m+n}(u_{m+n}),$$

where $u_i = \delta \exp(w_i)$ $(i = 1, \cdots, m+n)$. But, as is well known, $\delta \exp(w_i)$ is ordinary euclidean measure on the one (dimensional) linear space $\mathfrak{R}_i$ (cf.

[1], p. 155). Thus $\delta\theta(w_1 \times \cdots \times w_{m+n})$ is euclidean measure on $\mathfrak{R}^{m+n}$. Hence $G = E \cdot K(w_G; w_E, w_K)$ where w_G, w_K are left Haar measures, and w_E becomes ordinary euclidean measure when the coordinate system θ^{-1} is introduced on the analytic manifold E.

4. The theorems on embedding and on a base of analytic functions.

THEOREM 2. *A Lie group can be embedded analytically (with non-vanishing Jacobian) in Euclidean space.*

Proof. The Peter-Weyl Theorem tells us that a compact Lie group admits a (continuous) faithful finite dimensional representation. Since a continuous isomorphism of a Lie group is an analytic isomorphism ([1], p. 128), and since an analytic subgroup of the group of $n \times n$ matrices is an analytic submanifold of the euclidean space of all $n \times n$ matrices ([1], p. 101), a continuous faithful finite dimensional representation of a Lie group provides an analytic embedding with non-vanishing Jacobian in euclidean space. Consequently any analytic group G can be analytically embedded in euclidean space. For if $G = E \cdot K(\mathrm{AN})$ where K is compact and E is euclidean, and if θ is a faithful n-dimensional representation of K, then the mapping $\Theta:(e,k) \to (e, \theta(k))$ of G into the euclidean space $E \times E^{n^2}$ is an analytic isomorphism of G. Finally if G' is a Lie group, and if G_0 denotes the component of the identity of G, and if $\{G_n \mid n = 1, 2, \cdots\}$ denote the other connected components of G', then G' can be embedded analytically in Euclidean space as follows: Let θ be any analytic embedding of G_0 in the euclidean space E. Select for each G_i an element x_i $(i = 1, 2, \cdots)$. Define $\theta'(x_n g) = (\theta(g), n)$ for all $g \in G_0$ and for all n. Then θ' is an analytic embedding of G into the euclidean space $E_0 \times R$, where R denotes the real numbers.

THEOREM 3. *The L^2 space of a Lie group G has a base of functions which are analytic on G.*

Proof. We may obviously assume that G is connected. Thus $G = E \cdot K$ (W_G, W_E, W_K), where K is a compact subgroup, E is mapped by an analytic isomorphism θ onto a euclidean space R, and $\delta\theta(W_E)$ is ordinary euclidean measure on R. Let ϕ denote a faithful representation of the compact group K by matrices of degree n, let F denote the euclidean space of all $n \times n$ matrices (f_{ij}), the f_{ij} being regarded as coordinate functions $(i, j = 1, 2, \cdots, n)$. Let f'_{ij} denote the function $f_{ij}\phi$ and let $f'_1, f'_2, \cdots, f'_n, \cdots$ be an ortho-

normal base with respect to Haar measure on K of the ring of functions generated by $\{f'_{ij} \mid i, j = 1, \cdots, n\}$. Inasmuch as ϕ is an analytic *isomorphism* into F (i. e. with non-vanishing Jacobian), $f'_1, f'_2, \cdots$ are analytic on K and form, in fact, a base for the L^2 space of K (with respect to Haar measure). Let $\{g'_1, g'_2, \cdots, g'_n, \cdots\}$ be a base of analytic functions for the L^2 space of the euclidean space R. Let α, β denote the projections of G onto E, K respectively, and let $f_i = f'_i\beta$, $g_i = g'_i\theta\alpha$. Then $\{f_i \cdot g_j \mid i, j = 1, 2, \cdots\}$ is a base of analytic functions for the L^2-space of G.

SYRACUSE UNIVERSITY.

BIBLIOGRAPHY.

[1] C. Chevalley, *Theory of Lie Groups*, Princeton (1946).

[2] K. Iwassawa, "On some types of topological groups," *Annals of Mathematics*, vol. 50 (1949), pp. 507-558.

[3] A. Malcev, "On the theory of Lie groups in the large," *Matematichesk Sbornik*, vol. 58 (1945).

[4] G. D. Mostow, "A new proof of E. Cartan's theorem on the topology of semi-simple groups," *Bulletin of the American Mathematical Society*, vol. 55 (1949), pp. 969-980.

[5] L. Pontriagin, *Topological Groups*, Princeton (1940).

LOCALLY COMPACT RINGS. III.*

By Irving Kaplansky.

1. Introduction. In a series of papers ([3], [4], [5]) the author has carried out various investigations on locally compact rings. In the present paper we shall add three further theorems, as follows.

(a) Two earlier results ([4], theorems 13 and 14) are generalized by showing that a non-discrete locally compact primitive ring of characteristic 0 is a finite-dimensional algebra.

(b) A conjecture ([4], p. 459) is verified by showing that a non-discrete locally compact simple[1] ring with minimal ideals is a finite-dimensional algebra.

(c) A structure theory is given for locally compact right bounded semi-simple rings.

2. Primitive rings of characteristic 0. The theorem to be proved in this section is the following:

Theorem 1. *Let A be a non-discrete locally compact primitive ring of characteristic* 0. *Then A is a finite-dimensional algebra over its center.*

We need three lemmas. The first two are implicit in [4], Lemma 10 and Theorem 11, and are given without further proof.

Lemma 1. *Let A be a locally compact non-discrete totally disconnected ring. Then for some prime p, the set of all x with $p^n x \to 0$ forms a non-zero closed ideal in A.*

Lemma 2. *Let A be a locally compact ring which is not a Q-ring. Then there exists a non-zero idempotent e in A such that eAe is a Q-ring.*

The next lemma is a recapitulation of Jacobson's theory ([1], p. 236) of the eigenring; we drop the assumption of a unit element and insert topological trimmings.

* Received January 12, 1952.

[1] By a simple ring we shall always mean one that has no proper two-sided ideals, and not merely no proper closed two-sided ideals.

Lemma 3. *Let A be a topological ring and M a closed regular maximal right ideal in A. Let B be the set of all x in A with $xM \subset M$. Then B is a closed subring of A, M is a two-sided ideal in B, B/M is a topological division ring, and A/M is a topological linear space over B/M.*

Proof. Let e be a left unit mod M. The ring A acts by right multiplication on A/M to produce an irreducible ring of endomorphisms, say with commuting division ring D. Take θ in D and suppose θ sends the coset $e + M$ into $b + M$. Then since θ commutes with the right multiplication by x, we find that θ sends $ex + M$ into $bx + M$. Since $ex - x \in M$, we have $ex + M = x + M$. Thus θ sends $x + M$ into $bx + M$, and coincides with left multiplication by b. Necessarily $bM \subset M$, that is, $b \in B$. In this way we obtain a homomorphism of B onto D. The kernel K consists of all y in B with $yA \subset M$. Evidently K is a right ideal containing M. If $K \neq M$, then $K = A$, $A^2 \subset M$, which is impossible since e^2 is not in M. Thus $K = M$ and $B/M \cong D$. The remaining statements of the lemma are routine, and are left to the reader.

Proof of Theorem 1. *Case* I. A is a Q-ring and primary (in the sense that for some prime p, $p^n x \to 0$ for every x). Take a regular maximal right ideal M such that A is faithfully represented on A/M. Since A is a Q-ring, M is closed. So we are able to apply Lemma 3. The vital thing is to know that the locally compact division ring B/M is non-discrete; this follows from the assumption that A, and hence B/M, is primary. By [4], Lemma 9, A/M is finite-dimensional over B/M. From this it follows that A is a simple algebra finite-dimensional over its center.

Case II. A is primary but not a Q-ring. We apply Lemma 2 and obtain a non-zero idempotent e such that eAe is a Q-ring. Moreover eAe is again primitive, and again primary. By Case I, eAe is a finite-dimensional algebra. From this it follows that A at least has minimal ideals (indeed if f is a primitive idempotent in eAe, it is also a primitive idempotent in A). This allows us to quote [4], Theorem 14 to complete the proof of Case II.

Case III. A is arbitrary. By [4], Theorem 2 we can assume that A is totally disconnected. We cite Lemma 1, and find a non-zero closed primary ideal I in A. Also I is again primitive and again locally compact. On applying Case I or II, which ever is appropriate, we deduce that I is a finite-dimensional algebra. In particular it has a unit element, whence it is a direct sumamnd of A. This is incompatible with the primitivity of A, unless $I = A$. This concludes the proof of Theorem 1.

3. Simple rings with minimal ideals. With Theorem 1 we have concluded the study of locally compact primitive rings of characteristic 0. But characteristic p is another matter; the example on page 459 of [4] is an indication of the numerous possibilities. But if we go so far as to assume a simple ring with minimal ideals, we get the usual strong result.

THEOREM 2. *A non-discrete locally compact simple ring with minimal ideals is a finite-dimensional algebra over its center.*

Two lemmas are needed, the first purely algebraic.

LEMMA 4. *Let A be a simple ring with minimal ideals, and $\{e_i\}$ an infinite set of orthogonal idempotents in A. Then there cannot exist in A an element y such that, for every i, e_iy is a non-zero element of e_iAe_i.*

Proof. Represent A as a ring of finite-valued linear transformations on a vector space. Then the proposed element y, as a linear transformation, would have infinite-dimensional range. For its range contains the range (say S_i) of $e_iy = e_iye_i$; and each S_i is non-zero and is disjoint from the union of the remaining S_j's.

LEMMA 5. *Let A be a simple non-discrete topological ring, U a neighborhood of 0 in A, e an idempotent such that $U \cap (1-e)A(1-e) = 0$. Then e is the unit element of A.*

Proof. Suppose $e \neq 1$. By the simplicity of A we have $A = A(1-e)A$, whence

$$e = \sum_{i=1}^{n} x_i(1-e)y_i. \tag{1}$$

There exists a neighborhood V of 0 such that $V \subset U$ and

(2) $(1-e)V(1-e) \subset U$, (3) $(1-e)Vx_j(1-e) \subset U$,

(4) $(1-e)y_iV(1-e) \subset U$, (5) $(1-e)y_iVx_j(1-e) \subset U$.

Since $U \cap (1-e)A(1-e) = 0$, the left side of (2)-(5) is in each case 0. Multiply (3) on the right by y_j, add for $j = 1, \cdots, n$ and use (1); we get $(1-e)Ve = 0$. Similarly from (4) and (5) we find $eV(1-e) = 0$, $eVe = 0$. On adding these four equations we obtain $V = 0$, contradicting the non-discreteness of A.

Proof of Theorem 2. It is sufficient to prove that A has a unit element, for then it has the descending chain condition, etc. Suppose the contrary and let U be be a compact open subring in A. Take any non-zero element x_1 in U.

It is known ([2], Theorem 9) that $x_1 \varepsilon e_1Ae_1$ for a suitable idempotent e_1. By Lemma 5, we can pick a non-zero element x_2 in $U \cap (1-e_1)A(1-e_1)$; again $x_2 \varepsilon e_2Ae_2$, where e_2 is a suitable idempotent orthogonal to e_1. Continuing in this fashion, we construct sequences x_i, e_i with $x_i \neq 0$, $\{e_i\}$ orthogonal idempotents, $x_i \varepsilon U \cap e_iAe_i$. Write $y_i = x_1 + \cdots + x_i$; then $y_i \varepsilon U$. By compactness there exists in U a limit point y of the sequence $\{y_i\}$. We have $e_my_n = x_m$ for all $n \geqq m$, and hence $e_my = x_m$. We have contradicted Lemma 4.

4. Right bounded rings. In [4], Theorem 4 it was shown that a locally compact bounded semi-simple ring is a direct sum of a compact ring and a discrete ring. We shall now obtain a fairly complete structure theorem under the weaker hypothesis of right boundedness.

THEOREM 3. *Let A be a locally compact right bounded semi-simple ring. Then A contains an open ideal which is a local direct sum of finite simple rings relative to non-zero right ideals.*

LEMMA 6. *Let A be a semi-simple ring and I a right ideal in A. Then the radical $R(I)$ of I is its left annihilator. If $I/R(I)$ has a unit element, then I is of the form $I = eA$, e an idempotent.*

Proof. Let $x \varepsilon R(I)$, $y \varepsilon I$. To prove $xy = 0$ it is enough to prove that xy is in the radical of A; that is, we must show that xyz is quasi-regular for every z in A. But $x(yz) \varepsilon xI$ is in $R(I)$, and so is even quasi-regular in I.

Suppose further that $I/R(I)$ has a unit element, and let $e \varepsilon I$ be any element mapping on it. Then $(1-e)I \subset R(I)$, whence $[(1-e)I]^2 = 0$, $(1-e)I$ is a nilpotent right ideal in A, $(1-e)I = 0$. This proves that e is an idempotent and $I = eI \subset eA$. Further $I \supset eA$ and so $I = eA$.

LEMMA 7. *Let A be a semi-simple topological ring, and g a primitive idempotent in A. Then the closure of AgA is a primitive ring.*

Proof. Let C be the closure of AgA. We claim that the representation of C upon the right ideal $gC = gA$ is faithful. If not, there exists a non-zero element x in C with $gAx = 0$. But then $Cx = 0$, which contradicts the fact that C, an ideal in a semi-simple ring, is itself semi-simple.

LEMMA 8. *Let A be a locally compact right bounded semi-simple ring. Then A has a compact open right ideal of the form eA, with e an idempotent. Suppose further that P is a closed primitive ideal in A not containing e; then A/P is finite, and P is a direct summand of A.*

LEMMA 9. *A locally compact right bounded primitive ring is discrete.*

Proof. It is convenient to prove these two lemmas together. Let A be as in Lemma 8. An application of [4], Theorem 1 shows that A must be totally disconnected. Then A has compact open subgroups. By [3], Lemma 9 it even has a compact open right ideal I. Let $R(I)$ be the radical of I; $R(I)$ is closed by [5], Theorem 1. Then by [3], Theorem 16 $I/R(I)$ has a unit element. This makes Lemma 6 applicable and shows that I is of the form eA. We have proved the first part of Lemma 8.

We now suppose further that P is a closed primitive ideal in A not containing e. Write $B = A/P$, and let f be the image of e in B; note that f is non-zero. We have that fB is compact open, since it is the image of eA. Likewise fBf is compact, and moreover fBf (like B) is primitive. But a compact primitive ring is necessarily finite ([3] remark 7 on page 165). So fBf is finite, and this proves that B is a primitive ring with minimal ideals. Indeed f can be written as the sum of a finite number of idempotents which are primitive idempotents in fBf and likewise in B. Hence f lies in the unique minimal two-sided ideal of B, which we shall call D. Since D contains fB, it is open in B. Thus D is a locally compact simple ring with minimal ideals. If D is non-discrete, then by Theorem 2 it is a finite-dimensional algebra over a non-discrete center; but this is incompatible with the fact that D contains a compact open right ideal $fB = fD$. Hence D is discrete, whence fB is finite, from which it follows that B itself is finite (for an infinite primitive ring cannot have a finite non-zero right ideal).

We have now completed the proof of Lemma 8, except for the fact that P is a direct sumand of A. Before showing this, we pause to prove Lemma 9. We accordingly let A be a locally compact right bounded primitive ring and let eA be a compact open right ideal, as above. There are two possibilities. If $e = 0$, then A is of course discrete. If $e \neq 0$, we may take $P = 0$ in the above discussion, and we conclude that A is even finite.

We are now ready to conclude the proof of Lemma 8 by showing that P is a direct summand of A. For this purpose we look next at the compact semi-simple ring eAe, whose structure is completely known from [3], Theorem 16; in particular, it is well stocked with primitive idempotents, which are primitive idempotents also in A. It is not possible that P contains all these primitive idempotents; for then since it is closed it would contain e, the latter being a limit of sums of primitive idempotents. We pick a primitive idempotent g not in P. Let C be the closure of AgA. By Lemma

7, C is primitive. It is again right bounded,[2] and so Lemma 9 applies to show that C is discrete. That is, C is actually AgA and is a simple ring with minimal ideals. Since C is not in P, while both C and A/P are simple, we have a direct sum composition $A = C \oplus P$ at least in the algebraic sense. But this shows that C is isomorphic to A/P, hence is a finite simple ring with unit element; thus the direct sum is topological as well as algebraic. This completes the proof of Lemma 8.

Proof of Theorem 3. With Lemma 8 at hand, it is not difficult to analyze the structure of a locally compact right bounded semi-simple ring A. The basic tool is the compact open right ideal eA. We consider the closed primitive ideals in A; it follows from [5, Theorem 1] that their intersection is 0. Let J denote the intersection of those primitive ideals (necessarily open) that contain e. Since J contains eA, it is again open. We shall prove that J has the structure described in Theorem 3.

Let P_i denote a typical closed primitive ideal not containing e. By Lemma 8 we have $A = h_iA \oplus P_i$, where h_iA is a finite simple ring with unit element h_i. We observe that h_i lies in J. For let Q be a primitive ideal containing e. Then Q must contain either h_iA or P_i; but if it contains P_i it must equal P_i, since A/P_i is simple. This is a contradiction, and so Q contains h_iA. Thus h_i lies in J, which is the intersection of the Q's.

We define a mapping T of J into the complete direct sum of the rings $\{h_iA\}$ by sending $x \in J$ into $\{h_ix\}$. In the first place, T is faithful. For if $xT = 0$ then $h_ix = 0$, x lies in every P_i; while $x \in J$ implies that x lies in every remaining primitive ideal. Thus $x = 0$. Next the range of T includes at least the algebraic direct sum of the rings h_iA, that is, the elements with all but a finite number of coordinates zero; this is an automatic consequence of the fact that $\{h_iA\}$ are distinct minimal two sided ideals in J. From the compactness of eA we then deduce that eA is (algebraically and topologically) the complete direct sum of the right ideals h_ieA. Incidentally h_ie is non-zero, since otherwise e would lie in P_i.

At this point we know that J contains the local direct sum of the rings h_iA relative to the right ideals h_ieA. To complete the proof that J is precisely this local direct sum, only one thing more is needed; for any $y \in J$ we must show that all but a finite number of the elements h_iy lie in h_ieA. Let U be a neighborhood of 0 (we can suppose that $U \subset eA$) such that $yU \subset eA$. Since eA is precisely the Cartesian product of h_ieA, it follows that

[2] Right boundedness is inherited by subrings.

U is unrestricted at all but a finite number of coordinates. That is, for all but a finite number of i we have $yh_ieA \subset h_ieA$, which implies $y \varepsilon h_ieA$. This completes the proof of Theorem 3.

The structure of J is thus completely determined, while A/J can be an arbitrary discrete semi-simple ring. We shall not attempt to study the ring extension problem that arises here, but it should at least be remarked that J need not be a direct summand of A. To get a counter-example, take J to be the local direct sum of finite simple rings relative to *proper* right ideals, and let A be the result of adjoining to J the unit element of the complete direct sum.

UNIVERSITY OF CHICAGO.

BIBLIOGRAPHY.

[1] N. Jacobson, "Structure theory of simple rings without finiteness assumptions," *Transactions of the American Mathematical Society*, vol. 57 (1945), pp. 228-245.

[2] N. Jacobson and C. E. Rickart, "Jordan homomorphisms of rings," *ibid.*, vol. 69 (1950), pp. 479-502.

[3] I. Kaplansky, "Topological rings," *American Journal of Mathematics*, vol. 69 (1947), pp. 153-183.

[4] ———, "Locally compact rings," *ibid.*, vol. 70 (1948), pp. 447-459.

[5] ———, "Locally compact rings. II," *ibid.*, vol. 73 (1951), pp. 20-24.

CARTESIAN PRODUCTS OF REALS.*

By SAMUEL KAPLAN.

Introduction. Cartesian products of reals PR_λ have been finding increasing use in the theory of topological linear spaces, essentially because of their intimate connection with the weak topology of such spaces. An investigation of the basic properties of these products appears, therefore, to be desirable.

The primary problem is to free these spaces once and for all from their dependence on the special axes used to define them. In the present paper, this is attacked in two directions. The first direction, to which we devote part I, is that of finding intrinsic properties which characterize cartesian products of reals. We obtain two such characterizations. One of these is Lefschetz' *linear compactness* ([8], page 78), which he defined for his "linearly topologized vector spaces." Lefschetz showed (*cf.* [4]) that a linearly compact space of his type is a cartesian product of reals, the reals being taken as discrete. Clearly linear compactness as a concept can be defined equally well for ordinary topological linear spaces, and what we show is that the same structure theorem is obtained, viz.

A convex topological linear space is linearly compact if and only if it is a cartesian product of reals.

The second characterization is a not unexpected one:

A convex topological linear space is a cartesian product of reals if and only if it is complete in its weak topology.

The second direction, pursued in part III, is that of finding conditions, in a given cartesian product of reals PR_λ, for a set of one-dimensional linear subspaces $\{R_\alpha'\}$ to be an alternative set of axes for PR_λ. A necessary and sufficient condition is given in Theorem 11. One result in terms of matrices is the following (Corollary 2, Theorem 16):

Let $\{x^{(\alpha)}\}$ be a set of elements of PR_λ. Then a necessary condition that the $x^{(\alpha)}$'s generate an alternative set of axes is that the matrix of the coordinates of the $x^{(\alpha)}$'s be column-finite.

* Received October 4, 1951; revised February 23, 1952.

In part III there are also given the answers to some immediate questions about closed linear subspaces of cartesian products of reals. For example, *every closed linear subspace has a complementary closed linear subspace*, or (Mackey) *the join of two closed linear subspaces is closed.*

Part II is devoted to studying the conjugate space of a cartesian product of reals and using it to establish the reflexivity of the latter. The topology of the conjugate space turns out to be simply the supremum of all convex topologies which can be imposed on it.

In closing we wish to point out that the present paper is heavily indebted to Mackey's thesis ([10], [11]).

I. Characterization Theorems.

1. Basic definitions and propositions. All our spaces are real linear spaces, and we will omit the word "real." We use the word "linear"—as applied to both spaces and functions—in its purely algebraic sense.

We denote the identity of a linear space X by θ. If A and B are any two subsets of X, $A+B$ consists of all x of the form $a+b$, where $a \varepsilon A$, $b \varepsilon B$. For any collection of subsets $\{A_\lambda\}$, ΣA_λ will denote the set of all finite sums of the form $a_{\lambda_1}+\cdots+a_{\lambda_k}$, where $a_{\lambda_i} \varepsilon A_{\lambda_i}$ $(i=1,\cdots,k)$.

In a topological linear space, any open set containing θ will be called a *nucleus*. As is customary, we will give the topology of a topological linear space by simply giving a nuclear base.

Given a collection $\{X_\lambda\}$ of topological linear spaces, the *cartesian product* $X=PX_\lambda$ of the X_λ's is the space of all collections $x=\{x_\lambda\}$, where

(i) for each λ, the coordinate x_λ of x is an element of X_λ;

(ii) multiplication by reals and addition are coordinatewise;

(iii) we take for a nuclear base all sets of the form $\{x \mid x_{\lambda_i} \varepsilon U_{\lambda_i}; i=1,\cdots,k\}$, where $\lambda_1,\cdots,\lambda_k$ is any finite set of λ's, and U_λ is any nucleus in X_{λ_i} $(i=1,\cdots,k)$.

This topology is called the *Tychonoff*, or *product, topology*. We note that if the topology of every X_λ is convex, then that of X is also convex. For each λ_0, we identify X_{λ_0} with the subspace $\{x \mid x_\lambda=\theta \text{ for all } \lambda\neq\lambda_0\}$. As a result we have $X=(\Sigma X_\lambda)\dagger$ (the dagger means closure), any subproduct PX_μ $(\{\mu\}\subset\{\lambda\})$ can be written $PX_\mu=(\Sigma X_\mu)\dagger$, and a member of the nuclear base above can be written

$$U=\sum_{i=1}^{k} U_{\lambda_i}+(\sum_{\lambda\neq\lambda_i} X_\lambda)\dagger.$$

It is easily verified that if f is a continuous linear functional on PX_λ, then f is zero on all but a finite number of the X_λ's. It follows that every continuous linear functional on PX_λ can be written $f = f_{\lambda_1} + \cdots + f_{\lambda_k}$, where f_{λ_i} is a continuous linear function on X_{λ_i} $(i = 1, \cdots, k)$. And of course every such sum is a continuous linear functional on PX_λ.

If X is a topological linear space and X^* is the set of all continuous linear functionals on X, then the *weak topology* on X is defined as follows: For a nuclear base we take all sets of the form $\{x \mid |f_i(x)| < \epsilon \ (i = 1, \cdots, k)\}$, where $f_1, \cdots, f_k$ are any elements of X^*, and ϵ is any positive real number. It is easy to show that for $f_1, \cdots, f_k$ we need only consider linearly independent functionals.

If X is a linear space and F is the set of all linear functionals on X, then the *weak* topology* on F is defined as follows: For a nuclear base we take all sets of the form $\{f \mid |f(x_i)| < \epsilon \ (i = 1, \cdots, k)\}$, where $x_1, \cdots, x_k$ are any elements of X, and ϵ is any positive real number. Here also we need only consider linearly independent elements for $x_1, \cdots, x_k$.

We will need the following three (known) propositions.

(1) *Let X be a linear space, F the set of all linear functionals on X, and E any linear subspace of F which is total. Then under the weak* topology, $\bar{E} = F$.*

Proof. Consider any $f_0 \in F$. We have to show that given $\epsilon > 0$ and elements $x_1, \cdots, x_k$ of X, there exists an $f \in E$ such that $|f(x_i) - f_0(x_i)| < \epsilon$ for $i = 1, \cdots, k$. We will show in fact that there exists an $f \in E$ such that $f(x_i) = f_0(x_i)$ for $i = 1, \cdots, k$. If we let H be the subspace $\{f \in F \mid f(x_i) = 0 \ (i = 1, \cdots, k)\}$, then our last statement says that $f_0 + H$ contains elements of E; it is this that we will prove. We assume that $x_1, \cdots, x_k$ are linearly independent; then they are linearly independent as linear functionals on F. It follows ([3], Théorème 1) that F/H is exactly k-dimensional. Now E is total; hence $x_1, \cdots, x_k$ are also linearly independent as linear functionals on E. Therefore $E/H \cap E$ is also exactly k-dimensional, and so is identical with F/H. It follows that $F = E + H$, which immediately gives the required conclusion.

(2) *The set of all linear functionals on a linear space X forms, under the weak* topology, a space topologically isomorphic to a cartesian product of reals PR_ξ, where for the index set $\{\xi\}$, we can take any Hamel basis in X.*

Proof. Let $\{\xi\}$ be any Hamel basis for X, that is, every $x \in X$ can be written uniquely in the form $x = t_1\xi_1 + \cdots + t_k\xi_k$ (the t_i's real numbers).

If f is a linear functional on X, f is completely determined by the values $f(\xi)$, ξ running through the Hamel basis. Denoting $f(\xi)$ by f_ξ, the mapping $f \rightarrow \{f_\xi\}$ is clearly an algebraic isomorphism of the set of all linear functionals onto PR_ξ. It is easy to see that the Tychonoff topology on PR_ξ is precisely the weak* topology.

(3) *Let $X = PR_\xi$ be a cartesian product of reals. Then no convex topology on X different from the Tychonoff has exactly the same set of continuous linear functionals.*

Proof. As we remarked earlier, each continuous linear functional on PR_λ (under the Tychonoff topology) has the form $f = f_{\lambda_1} + \cdots + f_{\lambda_k}$, where f_{λ_i} is a linear functional on R_{λ_i} $(i = 1, \cdots, k)$. This implies immediately that the resulting weak topology is precisely the Tychonoff topology. Now suppose we are given a convex topology on PR_λ which has exactly the same set of continuous linear functionals as the Tychonoff. From what we have just remarked, this topology must be at least as fine as the Tychonoff. We show it cannot be finer.

Let U be any symmetric convex nucleus in this topology. We will show that U contains all but a finite number of the R_λ's, whence it will follow easily that U contains a Tychonoff nucleus. Let $p(x)$ denote the pseudo-norm defined on PR_λ by U in the usual manner ([12], § 9). What we have to prove is that $p(x)$ is zero on all but a finite number of the R_λ's. Suppose not. Then there is a denumerable set $\{R_{\lambda_n}\}$ on which $p(x)$ is different from zero. By the Hahn-Banach theorem, for each R_{λ_n}, there is a linear function $f^{(n)}$ such that (i) $f^{(n)}$ has norm 1 relative to $p(x)$; (ii) $f^{(n)}$ attains its norm on R_{λ_n}. From (i), each $f^{(n)}$ is bounded relative to $p(x)$, hence continuous in the topology, and hence by assumption continuous in the Tychonoff topology. It follows that it is zero on all but a finite number of the R_λ's. As a result (by taking a subsequence if necessary) we can assume each $f^{(n)}$ equals zero on all R_{λ_m}, $m > n$. Now define $f = \sum_{n=1}^{\infty} (1/3^n) f^{(n)}$. f is bounded relative to $p(x)$ and hence, by the same argument as above, equals zero on all but a finite number of R_λ's. But a simple calculation, using (ii), shows that f differs from zero on all the R_{λ_n}'s. This gives a contradiction.

In Mackey's terminology (*cf.* [11]), the above theorem states that a cartesian product of reals is both relatively weak and relatively strong. The theorem could have been proved from general considerations of Mackey on ideals of pseudonorm sets. Katětov also has a proof ([6], Theorem (2.20)), but it seems to contain an error.

2. The characterization theorems. Given a linear space X, by a *linear variety* we mean any set of the form $x + X_0$, where X_0 is a linear subspace. A topological linear space is called *linearly compacct* if, given any collection of closed linear varieties with the finite intersection property, the entire collection has a non-empty intersection.

THEOREM 1. *Let X be a convex topological space. Then a necessary and sufficient condition that X be linearly compact is that it be topologically isomorphic to a cartesian product of reals.*

Proof. Assume X is linearly compact. Let X^* be the set of all continuous linear functionals on X, and $\tilde{X}$ the set of all linear functionals on X^*. By the usual identification, $X \subset \tilde{X}$.

We impose the weak* topology on $\tilde{X}$. Since the resulting topology on X (as a subspace of $\tilde{X}$) is its weak topology, it follows that a set closed in X under the topology of $\tilde{X}$ is closed in X under the original topology.

We show first that X is all of $\tilde{X}$. Consider any $\tilde{x} \varepsilon \tilde{X}$. Let $\{W_\alpha\}$ be the collection of linear subspace of $\tilde{X}$ defined as follows: for each α, W_α is the set of linear functionals on X^* which are zero on some finite numbers of points of X^*, and every such set appears as a W_α. The W_α's are each closed; therefore the linear varieties $\{\tilde{x} + W_\alpha\}$ are each closed, and their intersection is exactly $\tilde{x}$. From the proof of (1), the closed linear varieties of X $\{(\tilde{x} + W_\alpha) \cap X\}$ have the finite intersection property. Hence, since X is linearly compact, they have a non-empty intersection, which from the above must be $\tilde{x}$. Thus $\tilde{x} \varepsilon X$, and we have $X = \tilde{X}$.

Now from (2), X is topologically isomorphic to a cartesian product of reals. Also, X and $\tilde{X}$ have the same set of continuous linear functionals ([3], Théorème 2). It follows from (3) that the topology of X is identical with that of $\tilde{X}$.

The converse property, that a cartesian product of reals is linearly compact, is proved by the same type of argument as that used in the Tychonoff Theorem (*cf.* [8], Chap. II (27.2)).

THEOREM 2. *Let X be a convex topological linear space. Then a necessary and sufficient condition that X be topologically isomorphic to a cartesian product of reals is that X be complete under its weak topology.*

Proof. Assume X is complete under its weak topology, and let X^* and $\tilde{X}$ be the same as in the previous proof. Since the topology of X as a subset of $\tilde{X}$ is its weak topology, X is a complete subset of $\tilde{X}$ and therefore closed in $\tilde{X}$. But from (1), $X\dagger - \tilde{X}$, which gives us $X = \tilde{X}$. That the

topology of X is identical with that of X follows by the precise argument given in the previous proof.

The converse part of the theorem follows from the fact that a cartesian product of reals is complete, and its topology is its weak topology (*cf.* (3)).

We summarize the above results in one theorem (in connection with (c), *cf.* [11], Theorem 15).

THEOREM 3. *For a convex topological linear space X, the following four properties are equivalent:* (a) *X is linearly compact*; (b) *X is complete in its weak topology*; (c) *X is complete and its topology is identical with its weak topology*; (d) *X is topologically isomorphic to a cartesian product of reals.*

To disassociate the concept of cartesian product of reals from dependence on any special set of axes, we will call such a space an *entire topological linear space.*

II. Reflexivity Properties.

3. Weak products of reals. The *weak* or *combinatorial* (*cf.* [6]) *product* $Y = P^w Y_\lambda$ of a set of convex topological linear spaces is the subspace ΣY_λ of PY_λ (that is, the points with only a finite number of coordinates different from the identity), but topologized as follows: each collection $\{V_\lambda\}$, where V_λ is a convex nucleus of Y_λ, determines a convex set V in Y, viz. the convex envelope of $\bigcup V_\lambda$; the set of all V's obtained in this manner is taken as a nuclear base for the topology of Y.

Note that we have defined the weak product of convex spaces only. This is merely for convenience, since we consider only convex spaces in this paper. It is possible to define the weak product of any topological linear spaces.

What we are interested in is the weak product of reals, $Y = P^w S_\lambda$ (throughout the paper we will denote the images of reals which occur in cartesian products by R's and those which occur in weak products by S's). We will aproach this by means of an intrinsic definition. That is, we define a certain type of convex topological linear space, study a few of its properties, and then show that it is topologically isomorphic to a weak product of reals.

Let Y be a linear space. We take for a nuclear base the collection $\{V\}$ of *all* sets containing θ with the following two properties: 1°. V is convex; 2°. V intersects every one-dimensional linear subspace in an open interval. The resulting topology will be called the *totally fine* convex linear topology on Y.

That $\{V\}$ does satisfy the conditions for a nuclear base is easily shown. The nuclear base defining any convex topology in Y always has properties 1° and 2°; hence the totally fine topology is clearly the finest convex topology on Y, the supremum of all the convex topologies ([2], page 11). We will call a linear space under its totally fine topology a *totally fine convex topological linear space* or simply a *totally fine space.* As the following theorems show, the properties with which we will be principally concerned in such a space are determined entirely by its algebraic structure.

THEOREM 4. *If Y is a totally fine space, every linear functional on Y is continuous.*

Proof. The totally fine topology, being the supremum of all convex topologies on Y, is at least as fine as the weak topology defined by the set of all linear functionals of Y. (It is actually much finer.)

THEOREM 5. *In a totally fine space, every linear subspace is closed.*

From Theorem 4, every hyperplane is closed. Hence a linear subspace, being the intersection of all the hyperplanes containing it, is closed.

By a Hamel basis of a linear space Y, we (for the moment) mean, as usual, a set $\{\eta_\lambda\}$ such that every $y \in Y$ is a unique finite linear combination of η_λ's. For each λ, let S_λ be the one-dimensional linear subspace generated by η_λ. We will find it more convenient to work with the S_λ's than with the η_λ's; hence we will adopt the convention of calling the collection $\{S_\lambda\}$ also a Hamel basis. Explicitly, a collection of one dimensional linear subspaces $\{S_\lambda\}$ is a Hamel basis if $Y = \Sigma S_\lambda$ and, for every λ_0, $S_{\lambda_0} \cap \sum_{\lambda \neq \lambda_0} S_\lambda = \theta$.

Now let Y be a linear space and $\{S_\lambda\}$ a Hamel basis for Y. Every collection $\{I_\lambda\}$, where I_λ is an open interval of S_λ containing θ, determines a convex set V' in Y, viz. the convex envelope of $\bigcup I_\lambda$. Let us call any set V' obtained in this way an $\{S_\lambda\}$-convex-set. It is easy to show that the collection of all $\{S_\lambda\}$-convex-sets forms a nuclear base for a convex topology on Y.

THEOREM 6. *Let Y be a linear space and $\{S_\lambda\}$ a Hamel basis for Y. The topology defined by the collection of all $\{S_\lambda\}$-convex-sets is identical with the totally fine topology.*

It is enough to show that every set satisfying 1° and 2° contains an $\{S_\lambda\}$-convex-set. Suppose V satisfies 1° and 2°. For each λ, let $I_\lambda = V \cap S_\lambda$. Since V is convex, the convex envelope V' of $\bigcup I_\lambda$ lies in V.

Given any Hamel basis $\{S_\lambda\}$ in a totally fine space Y, we may also consider the weak product $P^w S_\lambda$. Clearly, under the identification of each S_λ

of the Hamel basis with the corresponding S_λ of $P^w S_\lambda$, we have an algebraic isomorphism of Y with $P^w S_\lambda$. Theorem 6 shows that this is also a topological isomorphism. Thus we have the

COROLLARY. *A totally fine space is topologically isomorphic to the weak product of any Hamel basis in it.*

Since the S_λ's defining a weak product of reals obviously form a Hamel basis for the weak product, it also follows from Theorem 6 that every weak product of reals is a totally fine space. We thus have

THEOREM 7. *For a convex topological linear space Y, the following two properties are equivalent:* (a) *Y is totally fine*; (b) *Y is topologically isomorphic to a weak product of reals.*

Also of interest is

THEOREM 8. *A totally fine space is complete.*

The proof is given in the Appendix (§ 10).

4. Reflexivity. In normed linear spaces, the conjugate space of a given space X is topologized essentially by means of the bounded sets of X. We do the same here.

Given a topological linear space X, we topologize the set X^* of all continuous linear functionals on X as follows. For every closed bounded set G of X, let

$$N(B) = \{f \in X^* \mid \sup_{x \in B} |f(x)| < 1\}. \tag{4}$$

(Note that $N(B)$ is automatically convex.) We take for a nuclear base in X^* the collection of all $N(B)$'s. X^* topologized in this fashion will be called the *conjugate* space of X, and will be denoted by the same symbol X^*.

We will need the following two lemmas.

LEMMA 1. *Let $X = PX_\lambda$ be a cartesian product of linear topological spaces. Then every bounded set B of X is contained in a bounded "cube," that is, a set of the form $(\Sigma B_\lambda)\dagger$, B_λ a bounded set in X_λ.*

Proof. For each λ_0, let $B_{\lambda_0} = \{x_{\lambda_0} \mid x \in B\}$. We show B_{λ_0} is bounded in X_{λ_0}. Consider any nucleus U_{λ_0} of X_{λ_0}. Since B is bounded, there exists a real number $t > 0$ such that the nucleus $t(U_{\lambda_0} + (\sum_{\lambda \neq \lambda_0} X_\lambda)\dagger)$ contains B. But then tU_{λ_0} contains B_{λ_0}, hence the latter is bounded. The set $(\Sigma B_\lambda)\dagger$ is then the required "cube."

LEMMA 2. *Let $Y = P^w Y_\lambda$ be a weak product of convex topological linear spaces. Then every bounded set B lies in a finite dimensional subspace $Y_{\lambda_1} + \cdots + Y_{\lambda_k}$.*

Proof. Suppose B lies in no such subspace. Then there exists a sequence $\{y^{(n)}\} \subset B$ and a sequence $\{\lambda_n\}$ such that $y_{\lambda_n}{}^{(n)} \neq \theta$ $(n = 1, 2, \cdots)$. For each λ_n, choose a convex nucleus V_{λ_n} such that $y_{\lambda_n}{}^{(n)} \notin n V_{\lambda_n}$; and for all other λ's, take $V_\lambda = Y_\lambda$. Then if V is the convex envelope of $\bigcup V_\lambda$, $B \not\subset tV$ for any $t > 0$.

In what follows, if y is a linear functional on a linear space X, we will sometimes use the notation xy or yx to denote $y(x)$.

We now proceed with the proof of reflexity.

LEMMA 3. *Let $\{X_\lambda\}$ be a set of topological linear spaces. Then*

$$(PX_\lambda)^* = P^w X_\lambda{}^*.$$

Proof. For each λ, let us write $X_\lambda{}^* = Y_\lambda$. That $P^w Y_\lambda$ is the set of continuous linear functionals on PX_λ, under the definition $xy = \Sigma\, x_\lambda y_\lambda$, is proved in ([6], Theorem (2.2)). (It also follows from a remark we made in § 1). To establish the lemma, it is enough, from Lemma 1, to show that if B is of the form $B = (\Sigma B_\lambda)\dagger$ in PX_λ, then $N(B)$ is exactly the convex envelope of $\bigcup V_\lambda$, where $V_\lambda = N(B_\lambda)$ in Y_λ. This is readily verified by calculation.

LEMMA 4. *Let $\{Y_\lambda\}$ be a set of convex topological linear spaces. Then $(P^w Y_\lambda)^* = PY_\lambda{}^*$.*

Proof. For each λ, let us write $Y_\lambda{}^* = X_\lambda$. That PX_λ is the set of continuous linear functionals on $P^w Y_\lambda$ (with $xy = \Sigma\, x_\lambda y_\lambda$) is easily proved (*cf.* [6], Theorem 2.2). To establish the lemma it is enough, from Lemma 2, to show that if B is a bounded set in $Y_{\lambda_1} + \cdots + Y_{\lambda_k}$, then $N(B)$ in PX_λ is exactly the set $U(\lambda_1, \cdots, \lambda_k) + (\sum_{\lambda \neq \lambda_i} X_\lambda)\dagger$, where $U(\lambda_1, \cdots, \lambda_k) = N(B)$ in $X_{\lambda_1} + \cdots + X_{\lambda_k}$. Again this is readily verified by calculation.

If a topological linear space X has the property that $(X^*)^* = X$ (not only set-theoretically, but topologically) we will say that X is *reflexive.* (In this case X is convex). From Lemmas 3 and 4 we have immediately

LEMMA 5. *If $X = PX_\lambda$ and, for each λ, X_λ is reflexive, then X is also. If $Y = P^w Y_\lambda$ and, for each λ, Y_λ is reflexive, then Y is also.*

From this we conclude

THEOREM 9. *An entire topological linear space (a totally fine space) is reflexive, and its conjugate space is a totally fine space (an entire topological linear space).*

COROLLARY. *The converse of (2) holds, that is, an entire topological linear space is the set of all linear functionals on a linear space topologized by the weak* topology.*

The concept of conjugacy and reflexiveness can be formulated with closed bounded sets replaced by compact sets in (4). Lemmas 1 through 5 and Theorem 9 are easily seen to hold in this case too. Indeed we have a stronger conclusion. In the reals, the compact sets are identical with the closed bounded sets. It is easy to show, using Lemmas 1 and 2 that the same is true for cartesian and weak products of reals, that is

LEMMA 6. *In entire topological linear spaces and in totally fine spaces, the compact sets are identical with the closed bounded sets.*

III. Transformations of Axes.

5. Sets of axes. Let X be a linear space. Two linear subspaces X_1, X_2 are called *complementary* if $X_1 + X_2 = X$ and $X_1 \cap X_2 = \theta$. In such case it is trivial that every $x \varepsilon X$ can be written $x = x_1 + x_2$, $x_1 \varepsilon X_1$, $x_2 \varepsilon X_2$ in only one way.

Definition. Let X be a topological linear space. A set $\{R_\lambda\}$ of one-dimensional linear subspaces of X will be called *a set of axes for X* if properties (I) and (II) below hold.

(I) *For each partition of the index set $\{\lambda\}$ into two disjoint sets, $\{\lambda\} = \{\mu\} \cup \{\nu\}$, the closed linear subspaces $(\Sigma R_\mu)\dagger$ and $(\Sigma R_\nu)\dagger$ are complementary.*

Consider any $x \varepsilon X$. For every λ_0, R_{λ_0} and $(\sum_{\lambda \neq \lambda_0} R_\lambda)\dagger$ are complementary, by (I); hence x can be written uniquely $x = x_{\lambda_0} + x'$, $x_{\lambda_0} \varepsilon R_{\lambda_0}$, $x' \varepsilon (\sum_{\lambda \neq \lambda_0} R_\lambda)\dagger$. We will call x_{λ_0} the λ_0-*component* of x (with respect to the set $\{R_\lambda\}$). Now the finite subsets of the index set $\{\lambda\}$ are directed by inclusion, hence the finite sums $x_{\lambda_1} + \cdots + x_{\lambda_n}$ of the components of x form a directed set, and we can consider the question of the (Moore-Smith) convergence of this directed set. We can now state property (II).

(II) *For every $x \varepsilon X$, the directed set of all finite sums of the components of x converges to x. We denote this by $x = \Sigma x_\lambda$.*

We establish some immediate properties. In the following, $\{R_\lambda\}$ is a fixed set of axes.

(5) *For any two elements x, x' of X, $x = x'$ if and only if $x_\lambda = x'_\lambda$ for all λ.*

The necessity follows from the uniqueness of components; the sufficiency from the uniqueness of convergence.

(6) *If $\{\mu\}$ is a subset of $\{\lambda\}$, then a necessary and sufficient condition that $x \varepsilon (\Sigma R_\mu)\dagger$ is that $x_\lambda = \theta$ for all λ not in $\{\mu\}$.*

Proof. Assume $x \varepsilon (\Sigma R_\mu)\dagger$, and consider any λ_0 not in $\{\mu\}$. We can write $x = x + \theta$, where $x \varepsilon (\Sigma R_\mu)\dagger \subset (\sum_{\lambda \neq \lambda_0} R_\lambda)\dagger$ and $\theta \varepsilon R_{\lambda_0}$; hence from the uniqueness of components, $x_{\lambda_0} = \theta$. Conversely, suppose $x_\lambda = \theta$ for all λ not in $\{\mu\}$. Then $x = \Sigma x_\lambda = \Sigma x_\mu \varepsilon (\sum R_\mu)\dagger$.

(7) *Let $\{\mu\}$ be a subset of $\{\lambda\}$ and $x \varepsilon X$. We have $x = y + z$, where $y \varepsilon (\Sigma R_\mu)\dagger$ and $z \varepsilon (\sum_{\lambda \neq \mu} R_\lambda)\dagger$. Then* (i) *$y_\mu = x_\mu$ for all μ;* (ii) *$y_\lambda = \theta$ for all λ not in $\{\mu\}$.*

(ii) was proved in (6). We prove (i). For any μ_0, $y = y_{\mu_0} + y'$, $y' \varepsilon (\sum_{\lambda \neq \mu_0} R_\lambda)\dagger$. Hence $x = y_{\mu^0} + (y' + z)$, $y' + z \varepsilon (\sum_{\lambda \neq \mu_0} R_\lambda)\dagger$. It follows that $x_{\mu_0} = y_{\mu_0}$. (Note that the proof of (7) uses only property (I)).

(8) *Suppose that for each σ, Λ_σ is a subset of $\{\lambda\}$. If $\{\mu\} = \bigcap \Lambda_\sigma$, then $(\Sigma R_\mu)\dagger = \bigcap_\sigma (\sum_{\lambda \varepsilon \Lambda_\sigma} R_\lambda)\dagger$* (If $\bigcap \Lambda_\sigma$ is empty, we write $(\Sigma R_\mu)\dagger = \theta$).

Proof. The left side is clearly contained in the right. To show the inverse inclusion, consider $x \notin (\Sigma R_\mu)\dagger$. From (6), $x_{\lambda_0} \neq \theta$ for some λ_0 not in $\{\mu\}$. Choose σ such that $\lambda_0 \notin \Lambda_\sigma$. Then, from (6) again, $x \notin (\sum_{\lambda \varepsilon \Lambda_\sigma} R_\lambda)\dagger$, hence, a fortiori, x is not an element of the right side.

THEOREM 10. *In a totally fine space Y,* (a) *every Hamel basis is a set of axes;* (b) *every set of axes is a Hamel basis;* (c) *property* (I) *implies property* (II).

Proof. Let $\{S_\lambda\}$ be a Hamel basis. Then, for every partition of $\{\lambda\}$ into two disjoint sets, $\{\lambda\} = \{\mu\} \cup \{\nu\}$, ΣS_μ and ΣS_ν are complementary. But from Theorem 5, $\Sigma S_\mu = (\Sigma S_\mu)\dagger$ and $\Sigma S_\nu = (\Sigma S_\nu)\dagger$; therefore property

(I) holds. Since every $y \varepsilon Y$ lies in a finite sum of the S_λ's, property (II) is trivially satisfied. This proves (a).

Now suppose $\{S_\lambda\}$ is a set of axes. This means $Y = (\Sigma S_\lambda)\dagger = \Sigma S_\lambda$ and for each λ_0, $S_{\lambda_0} \cap (\sum_{\lambda \neq \lambda_0} S_\lambda)\dagger = \theta$; thus $\{S_\lambda\}$ is a Hamel basis. This proves (b). In proving (b), we have used only property (I), hence, applying (a), we have (c).

If X is a topological linear space and X_0 a linear subspace, then by $X_0{}^\perp$ we will mean the set of all continuous linear functionals which are zero on X_0. The following proposition is well known.

(9) *Let X and Y be convex topological linear spaces such that each is the set of continuous linear functionals on the other. Then*

(a) *if X_0 is a linear subspace of X, $(X_0{}^\perp)^\perp = \bar{X}_0$;*

(b) *if $\{X_\lambda\}$ is a collection of linear subspaces of X, then $(\Sigma X_\lambda)^\perp = \cap X_\lambda{}^\perp$;*

(c) *if $(X_1 + X_2)\dagger = X$, then $X_1{}^\perp \cap X_2{}^\perp = \theta$;*

(d) *if $\bar{X}_1 \cap \bar{X}_2 = \theta$, then $(X_1{}^\perp + X_2{}^\perp)\dagger = Y$.*

Suppose X and Y are convex topological linear spaces such that each is the set of continuous linear functionals of the other. Let $\{R_\lambda\}$ be a set of axes in X. Then a set of axes $\{S_\lambda\}$ in Y (with the same index system) will be said to be *orthogonal* to $\{R_\lambda\}$ if, for every λ_0,

$$S_{\lambda_0} = (\sum_{\lambda \neq \lambda_0} R_\lambda)^\perp \tag{10}$$

Note that there cannot be more than one set of axes in Y orthogonal to $\{R_\lambda\}$, since (10) deterimnes each S_{λ_0} uniquely.

(11) *Under the above conditions:*

(i) *For every partition of $\{\lambda\}$ into two disjoint sets,*

$$\{\lambda\} = \{\mu\} \cup \{\nu\}, \qquad (\Sigma S_\nu)\dagger = (\Sigma R_\mu)^\perp;$$

(ii) *$\{R_\lambda\}$ is in turn orthogonal to $\{S_\lambda\}$.*

Proof. From (9a), for each λ_0, $(\sum_{\lambda \neq \lambda_0} R_\lambda)\dagger = S_{\lambda_0}{}^\perp$. Hence, from (9b) and (8), $(\Sigma R_\mu)\dagger = (\Sigma S_\nu)^\perp$. Applying (9a) again gives (i). To establish (ii) we put $\{\mu\} = \lambda_0$ in (i) and obtain $(\sum_{\lambda \neq \lambda_0} S_\lambda)\dagger = R_{\lambda_0}{}^\perp$. Then another application of (9a) gives $R_{\lambda_0} = (\sum_{\lambda \neq \lambda_0} S_\lambda)\dagger^\perp$, which is (ii).

6. Axes in entire topological linear spaces. It is of course trivial that

LEMMA 7. *In a cartesian product of reals PR_λ, the collection $\{R_\lambda\}$ is a set of axes and is orthogonal to the collection $\{S_\lambda\}$ of the conjugate space $P^{w}S_\lambda$.*

The question arises: What other sets of axes are possible for PR_λ besides the set of R_λ's defining it? One answer is given by

THEOREM 11. *Let X be an entire topological linear space and Y its conjugate totally fine space. Then there is a one-one correspondence between the sets of axes in X and the Hamel bases in Y, the correspondence being given by orthogonality.*

Proof. Consider any Hamel basis $\{S_\lambda\}$ in Y. For each λ_0, let $R'_{\lambda_0} = (\sum_{\lambda\neq\lambda_0} S_\lambda)^\perp$. We show that the resulting collection $\{R'_\lambda\}$ is a set of axes in X. Now Y is topologically isomorphic to $P^{w}S_\lambda$ (Corollary to Theorem 6), and therefore its conjugate space is PR_λ. Hence, since X is conjugate to Y, there is determined a topological isomorphism of X on PR_λ. Moreover, since $(\sum_{\lambda\neq\lambda_0} S_\lambda)^\perp$ is R'_{λ_0} in X and R_{λ_0} in PR_λ, we must have $R'_{\lambda_0} \to R_{\lambda_0}$ under this isomorphism. It follows that $\{R'_\lambda\}$ is a set of axes in X.

Conversely, consider any set of axes $\{R_\lambda\}$ in X. For each λ_0, let $S_{\lambda_0} = (\sum_{\lambda\neq\lambda_0} R_\lambda)^\perp$. We show that $\{S_\lambda\}$ is a Hamel basis in Y. In the proof of (11) we did not use the fact that $\{S_\lambda\}$ was a set of axes, hence we can apply the results here. This gives first (using (9a)) that for each λ_0, $S_{\lambda_0} = (\sum_{\lambda\neq\lambda_0} R_\lambda)^{\dagger\perp}$, $\sum_{\lambda\neq\lambda_0} S_\lambda = R_{\lambda_0}{}^\perp$, hence (using (9c)) that $S_{\lambda_0} \cap \sum_{\lambda\neq\lambda_0} S_\lambda = \theta$. Secondly, it gives $\sum_\lambda S_\lambda = \theta^\perp = Y$, and thus $\{S_\lambda\}$ is a Hamel basis.

That the correspondence between the sets of axes in X and the Hamel bases in Y is one-one follows from the uniqueness of orthogonality. This completes the proof.

COROLLARY. *If $\{R_\lambda\}$ and $\{R'_\alpha\}$ are two sets of axes for an entire topological linear space, then the index sets $\{\lambda\}$ and $\{\alpha\}$ have the same cardinality.*

This follows from the fact that all Hamel bases in Y have the same cardinality ([8], Chap. II (24.1)).

From the proof of Theorem 11, we also have

THEOREM 12. *If X is an entire topological linear space and $\{R_\lambda\}$ a collection of one-dimensional linear subspaces, then X is topologically isomorphic to PR_λ with $R_\lambda \to R_\lambda$ if and only if $\{R_\lambda\}$ is a set of axes.*

COROLLARY. *If X is an entire topological linear space and $\{R_\lambda\}$ is a set of axes, then the sets of the form*

$$U(\lambda_1, \cdots, \lambda_k) + (\sum_{\lambda \neq \lambda_i} R_\lambda)\dagger, \tag{12}$$

where $U(\lambda_1, \cdots, \lambda_k)$ is a nucleus in the finite dimensional subspace $\sum_{i=1}^{k} R_{\lambda_i}$, form a nuclear base for X.

7. Subspaces of entire topological linear spaces.

THEOREM 13. *Let X be an entire topological linear space. If X_0 is a closed subspace, there exists a set of axes $\{R_\lambda\}$ such that $X_0 = (\Sigma R_\mu)\dagger$, where $\{\mu\}$ is a subset of $\{\lambda\}$.*

Proof. Let Y be the totally fine space conjugate to X. Choose a Hamel basis $\{S_\lambda\}$ for Y such that $X_0{}^\perp = \Sigma S_\nu$, where $\{\nu\}$ is a subset of $\{\lambda\}$. Denote the remaining λ's by $\{\mu\}$. Now let $\{R_\lambda\}$ be the set of axes in X orthogonal to $\{S_\lambda\}$. Then from (11), $(\Sigma R_\mu)\dagger = (\Sigma S_\nu)^\perp = X_0$.

If X is a topological linear space and X_0 is a linear subspace, a *projection* of X on X_0 is a continuous linear mapping E of X onto X_0 such that $E^2(X) = E(X)$ for all x. In such case, as is easily shown, $E^{-1}(\theta)$ is complementary to X_0. In a cartesian product PX_λ it is obvious that the natural mapping onto any subproduct PX_μ is a projection. Now in Theorem 13, X is topologically isomorphic to PR_λ, which gives us

THEOREM 14. *In an entire topological linear space, if X_0 is a closed linear subspace, there exists a projection on X_0 (and hence X_0 has a complementary closed linear subspace).*

It is also easily shown that

THEOREM 15. *If X is an entire topological linear space, and X_1, X_2 are complementary closed linear subspaces, then there exists a set of axes $\{R_\lambda\}$ such that $X_1 = (\Sigma R_\mu)\dagger$, $X_2 = (\Sigma R_\nu)\dagger$, where $\{\lambda\} = \{\mu\} \cup \{\nu\}$.*

In any topological linear space X, if X/X_0 is finite dimensional, the subspace X_0 is said to have *finite deficiency* ([10], Chap. I, § 1).

THEOREM 16. *Let X be an entire topological linear space and $\{R_\lambda\}$ a set of axes. Then if X_0 is any closed linear subspace of finite deficiency, X_0 contains all but a finite number of the R_λ's.*

Proof. Let $\{S_\lambda\}$ be the Hamel basis in the conjugate space Y which

is orthogonal to $\{R_\lambda\}$. $X_0{}^\perp$ is finite dimensional ([3], Théorème 1); hence there exixst $\lambda_1, \cdots, \lambda_k$ such that $X_0{}^\perp \subset \sum_{i=1}^{k} S_{\lambda_i}$. Since $(\sum_{\lambda \neq \lambda_i} R_\lambda)\dagger = (\sum_{i=1}^{k} S_{\lambda_i})^\perp$, it follows that $(\sum_{\lambda \neq \lambda_i} R_\lambda)\dagger \subset (X_0{}^\perp)^\perp = X_0$.

COROLLARY 1. *If $\{R_\lambda\}$, $\{R'_\alpha\}$ are two sets of axes for an entire topological linear space, then for any finite set $\lambda_1, \cdots, \lambda_k$, $(\sum_{\lambda \neq \lambda_i} R_\lambda)\dagger$ contains all but a finite number of the R'_α's.*

In terms of matrices, this can be stated as follows.

COROLLARY 2. *Given PR_λ, suppose $\{z^{(\alpha)}\}$ generates a set of axes (that is, the set of one-dimensional linear subspaces determined by the respective $z^{(\alpha)}$'s is a set of exes). Then the matrix of the coordinates of the $z^{(\alpha)}$'s is column-finite.*

Given two closed linear subspaces X_1, X_2 of a topological linear space two questions that arise are the following: "Is $X_1 + X_2$ closed?" and "If $X_1 \cap X_2 = \theta$, is $X_1 + X_2$ topologically isomorphic to the cartesian product of X_1 and X_2?" In general, the answer is no to both questions ([10], pp. 173-174). However Mackey has shown that in an entire topological linear space, the answer to both questions is yes. Because of its interest we state this as a theorem.

THEOREM 17. (*Mackey*). *In an entire topological linear space, if X_1, X_2 is any pair of closed linear subspaces, then* (a) *$X_1 + X_2$ is closed*; (b) *if $X_1 \cap X_2 = \theta$, then $X_1 + X_2$ is topologically isomorphic to the cartesian product of X_1 and X_2.*

The theorem follows from ([10], Corollary 3 of Theorem III-6, Theorem III-7, and the paragraph preceding Theorem III-6).

THEOREM 18. *If the index set $\{\lambda\}$ has cardinality $2^{\aleph_0}$, then $X = PR_\lambda$ contains a linear subspace of dimension $\aleph_0$ which is dense in X (thus X is separable).*

Proof. Since all cartesian products of the same number of images of reals are topologically isomorphic, it is enough to exhibit one example with the above property. Let $Z = P^w R'_n$ $(n = 1, 2, \cdots)$ considered simply algebraically. Let Y be the space of all linear functionals on Z. The cardinality of any Hamel basis in Y is clearly $2^{\aleph_0}$. Now let X be the space of all linear functionals on Y, with the weak* topology. From (2), X has the form PR_λ, where the cardinality of $\{\lambda\}$ is $2^{\aleph_0}$; and from (1), Z is dense in X.

It is obvious the above theorem holds with $\aleph_0$ replaced by any infinite cardinal.

Appendix.

8. Sets of axes in Banach spaces. An interesting question on sets of axes is the following: does property (II) follow from property (I)? We show here that such is the case in Banach spaces. In fact we have a stronger result:

THEOREM 19. *Let X be a Banach space and $\{R_\lambda\}$ a set of one-dimensional linear subspaces having property* (I). *Then for each $x \varepsilon X$, all but a countable number $\{x_n\}$ of the components are zero and* $\lim_{n\to\infty} \| x - \sum_{i=1}^{n} x_i \| = 0$.

To prove this we make use of a theorem proved independently by E. R. Lorch ([9], Theorems (2. 1) and (2. 2)) and H. Kober [7]. This theorem says essentially that in a Banach space, properties (a) and (b) of Theorem 17 are equivalent. However we state it in a form adapted to the present requirements.

(13) (*Lorch-Kober*) *Let X be a Banach space and X_1, X_2 two complementary closed subspaces. For every $x \varepsilon X$, let $x = x_1 + x_2$, $x_1 \varepsilon X_1$, $x_2 \varepsilon X_2$, be the unique representation of x with respect to X_1, X_2. Then the mapping of X onto X_1 given by $x \to x_1$ is continuous.*

We proceed with the proof of Theorem 19. Consider $x \varepsilon X$. For any subset $\{\mu\}$ of the index set $\{\lambda\}$, we have $x = x' + x''$, where $x' \varepsilon (\Sigma R_\mu)\dagger$ and $x'' \varepsilon (\sum_{\lambda \neq \{\mu\}} R_\lambda)\dagger$. We will call x' the *component* of x in $(\Sigma R_\mu)\dagger$. Now $x \varepsilon (\Sigma R_\lambda)\dagger$; therefore there exists $y^{(1)} \varepsilon \sum_{n=1}^{n_1} R_{\lambda_n}$ such that $\| x - y^{(1)} \| < 1$. Denote $\sum_{n=1}^{n_1} R_{\lambda_n}$ by Y_1 and the component of x in Y_1 by $x^{(1)}$. It follows that

$$(x - x^{(1)}) \varepsilon (\sum_{\lambda \neq \lambda_i} R_\lambda)\dagger;$$

hence there exist $y^{(2)} \varepsilon \sum_{n=n_1+1}^{n_2} R_{\lambda_n}$ such that $\|(x - x^{(1)}) - y^{(2)}\| < 1/2$. Denote $\sum_{n=n_1+1}^{n_2} R_{\lambda_n}$ by Y_2 and the component of x in Y_2 by $x^{(2)}$. Continuing in this fashion, we obtain sequences $\{Y_m\}$, $\{y^{(m)}\}$, $\{x^{(m)}\}$ such that

(i) $\quad x^{(m)}$ is the component of x in Y_m;

(ii) $$\lim_{M\to\infty} \| (x - \sum_{m=1}^{M-1} x^{(m)}) - y^{(M)} \| = 0.$$

We will show that $\lim_{M\to\infty} \| x - \sum_{m=1}^{M} x^{(m)} \| = 0$, which will prove the theorem (Cf. (6)). Now $\| x - \sum_{m=1}^{M} x^{(m)} \| \leqq \| (x - \sum_{m=1}^{M-1} x^{(m)}) - y^{(M)} \| + \| y^{(M)} - x^{(M)} \|$; hence from (ii), it is enough to show that $\lim_{M\to\infty} \| y^{(M)} - x^{(M)} \| = 0$.

Suppose this last is not true. Then there exists $\epsilon > 0$ and a subsequence $\{M_p\}$ such that

(iii) $\| y^{(M_p)} - x^{(M_p)} \| \geqq \epsilon$ for $p = 1, 2, \cdots,$

and by dropping out terms if necessary, we can assume:

(iv) No two successive M's are in the subsequence $\{M_p\}$.

Now from (ii), $\lim_{M\to\infty} \| (\sum_{m=1}^{M-1} x^{(m)} + y^{(M)}) - (\sum_{m=1}^{M} x^{(m)} - y^{(M+1)}) \| = 0$, and this reduces to

(v) $\lim_{M\to\infty} \| y^{(M)} - x^{(M)} - y^{(M+1)} \| = 0.$

Let $Z = (\Sigma X_{M_p})\dagger$. From (iv), the component of $y^{(M_p)} - x^{(M_p)} - y^{(M_p+1)}$ in Z is $y^{(M_p)} - x^{(M_p)}$. Hence from (v) and the Lorch-Kober Theorem above, $\lim_{p\to\infty} \| y^{(M_p)} - x^{(M_p)} \| = 0$, which contradicts (iii).

9. In contrast to the above, we now show that in an entire topological linear space, property (I) does *not* imply property (II). Let Z, Y, X be the spaces in the proof of Theorem 18. We show first that the set $\{R'_n\}$ has property (I) in X. Let $\{n\} = \{p\} \cup \{q\}$ be a partition of $\{n\}$. That $(\Sigma R'_p)\dagger + (\Sigma R'_q)\dagger = X$ follows from Theorem 17 and Theorem 18. We show that $(\Sigma R'_p)\dagger \cap (\Sigma R'_q)\dagger = \theta$ by showing that $(\Sigma R'_p)^{\perp} + (\Sigma R'_q)^{\perp} = Y$. Now Y can be considered as the space of all sequences, $y = (y_1, y_2, \cdots, y_n, \cdots)$, and clearly $(\Sigma R'_p)^{\perp}$ consists of exactly the sequences with zero in the p-positions while $(\Sigma R'_q)^{\perp}$ consists of exactly the sequences with zero in the q-positions. It is obvious that these two classes of sequences generate Y.

Thus the set R'_n has property (I). If it had property (II), it would constitute a set of axes, which would contradict the Corollary to Theorem 11. This completes the proof.

10. Completeness of totally fine spaces. We present here the proof of Theorem 8. From Theorem 7, it is enough to give the proof for a weak product of reals $Y = P^w S_\lambda$. By definition an element V of the nuclear base is the convex envelope of a set $\bigcup_\lambda I_\lambda$, where I_λ is an open interval of S_λ

containing zero. Clearly, we can confine ourselves to I_λ's which are symmetric around zero, and we shall do so. If we denote the half-length of each I_λ defining V by e_λ (e_λ may possibly be ∞), then V is completely determined by the e_λ's as follows:

$$V = \{y \mid \Sigma \mid y_\lambda \mid / e_\lambda < 1\}. \tag{14}$$

(Since only a finite number of the y_λ's in any one y are different from zero, the above sum is always well defined.)

Now consider a Cauchy system $\{y^{(\alpha)}\}$ in Y. This means that given any collection $\{e_\lambda\}$, there exists α such that

$$\text{(i)} \qquad \beta > \alpha \text{ implies } \Sigma \mid y_\lambda^{(\beta)} - y_\lambda^{(\alpha)} \mid / e_\lambda < 1.$$

This gives, in particular, that

$$\text{(ii)} \qquad \beta > \alpha \text{ implies } \mid y_\lambda^{(\beta)} - y_\lambda^{(\alpha)} \mid < e_\lambda \text{ for every } \lambda.$$

This latter implies that the coordinates of the $y^{(\alpha)}$'s converge on each S_λ. We denote the limit on each S_λ by y_λ.

We show first that $y_\lambda = 0$ for all but a finite number of λ's. Suppose not. Then for some denumerable set of λ's, $\{\lambda_n\}$, the y_{λ_n}'s are all different from zero. Choose $e_{\lambda_n} = \frac{1}{2} \mid y_{\lambda_n} \mid$ $(n = 1, 2, \cdots)$ and for the rest of the λ's, choose $e_\lambda = \infty$. Let α satisfy (ii) with respect to the set $\{e_\lambda\}$. Then since $y_\lambda^{(\alpha)} = 0$ for all but a finite number of λ's, there is an n such that $y_{\lambda_n}^{(\alpha)} = 0$. But then, for this n, $\beta > \alpha$ implies $y_{\lambda_n}^{(\beta)} < \frac{1}{2} \mid y_{\lambda_n} \mid$, which contradicts the fact that $\lim_\alpha y_{\lambda_n}^{(\alpha)} = y_{\lambda_n}$.

Since only a finite number of the y_λ's are thus different from zero, $y = \sum_\lambda y_\lambda$ is an element of Y. We complete the proof by showing that $\lim y^{(\alpha)} = y$. Given any set $\{e_\lambda\}$, choose α to satisfy (i). Let $\lambda_1, \cdots, \lambda_k$ be the set of indices for which the coordinates of either y or $y^{(\alpha)}$ are different from zero. Since $\sum_{i=1}^{k} S_{\lambda_i}$ is finite dimensional, we can find $\alpha' > \alpha$ such that

$$\text{(iii)} \qquad \beta > \alpha' \text{ implies } \sum_{i=1}^{k} \mid y_{\lambda_i}^{(\beta)} - y_{\lambda_i} \mid / e_{\lambda_i} < 1.$$

For the rest of the λ's we have (since $y_\lambda = 0 = y_\lambda^{(\alpha)}$)

$$\text{(iv)} \qquad \beta > \alpha' \text{ implies } \sum_{\lambda \neq \lambda_i} \mid y_\lambda^{(\beta)} - y_\lambda \mid / e_\lambda = \sum_{\lambda \neq \lambda_i} \mid y_\lambda^{(\beta)} \mid / e_\lambda$$

$$= \sum_{\lambda \neq \lambda_i} \mid y_\lambda^{(\beta)} - y_\lambda^{(\alpha)} \mid / e_\lambda \leqq \sum_{\text{all } \lambda} \mid y_\lambda^{(\beta)} - y_\lambda^{(\alpha)} \mid / e_\lambda < 1,$$

from (i). Combining (iii) and (iv), we obtain that

$$\beta > \alpha' \text{ implies } \sum_{\lambda} | y_\lambda^{(\beta)} - y_\lambda |/e_\lambda < 2.$$

This completes the proof.

Remark. It may be of interest to note that a weak product of a countably infinite set of reals $P^w S_n$ furnishes a simple example of a complete topological linear space which is of the first category.

As a final remark, we point out that the completeness of entire and totally topological linear spaces are, as is not hard to see, just special cases of the following.

If $\{X_\lambda\}$ is a set of complete topological linear spaces, then PX_λ and $P^w X_\lambda$ are also complete.

WAYNE UNIVERSITY.

BIBLIOGRAPHY.

[1] R. Arens, "Duality in linear spaces," *Duke Mathematical Journal*, vol. 14 (1947), pp. 787-794.

[2] N. Bourbaki, *Elements de Mathématique*, Livre III (Actualités Scientifiques et Industrielles, 1940).

[3] J. Dieudonné, "La dualité dans les espaces vectoriels topologiques," *Annales de l'École Normale Supérieure*, vol. LIX (1942), pp. 107-139.

[4] ———, "Linearly compact spaces and double vector spaces over sfields," *American Journal of Mathematics*, vol. 72 (1951), pp. 13-19.

[5] S. Kaplan, "Extensions of the Pontrjagin duality, I: Infinite products," *Duke Mathematical Journal*, vol. 15 (1948), pp. 649-658.

[6] M. Katětov, "On convex topological linear spaces," Acta Facultatis Rerum Naturalium Universitatis Carolinae, no. 181 (1948), pp. 3-20.

[7] H. Kober, "A theorem on Banach spaces," Compositio Mathematicae, vol. 7 (1939), pp. 135-140.

[8] S. Lefschetz, *Algebraic Topology*, American Mathematical Society Colloquium Publications, vol. XXVII, New York, 1942.

[9] E. R. Lorch, "On a calculus of operators in reflexive vector spaces," *Transactions of the American Mathematical Society*, vol. 45 (1939), pp. 217-234.

[10] G. W. Mackey, "On infinite-dimensional linear spaces," *ibid.*, vol. 57 (1945), pp. 155-207.

[11] ———, "On convex topological linear spaces," *ibid.*, vol. 60 (1946), pp. 519-537.

[12] J. von Neumann, "On complete topological spaces," *ibid.*, vol. 37 (1935), pp. 1-20.

NORMAL CURVATURE OF A VECTOR FIELD.*

By T. K. Pan.

1. **Introduction.** Let v be a vector field in a surface in an ordinary space. The derived vector of v along a curve C on the surface is called *the absolute curvature vector of the field with respect to the curve.* This derived vector can be decomposed into a component tangential and a component normal to the surface. The former, called *the associate curvature vector of v along C* or *the angular spread vector of v along C*, was studied by W. C. Graustein [1] and was later generalized to Riemannian space by R. M. Peters [2]. This paper deals with the latter component, which is called *the normal curvature vector of the vector field v with respect to the curve C.* The following terms are defined: normal curvature of a vector field with respect to a curve, asymptotic direction of a vector field, asymptotic line of a vector field, principal direction of a vector field, line of curvature of a vector field, principal curvature of a vector field, indicatrix of a vector field, curve of a vector field, and principal vector field of the surface; new characterizations of the principal curvatures of the surface and the lines of curvature of the surface are derived; and some properties analogous to those connected with normal curvature of a curve on the surface are studied, first for vector fields in a surface in an ordinary space, and then extended to fields in V_n in V_{n+1} and to fields in V_n in V_m.

The notation of Eisenhart [3], [4] will be used for the most part except that $\Gamma^\alpha_{\beta\gamma}$ will be employed for Christoffel symbols of the second kind.

2. **Definitions.** Let $S: x^i = x^i(u^1, u^2)$, $i = 1, 2, 3$ be a real proper analytic surface in an ordinary space with reference to a rectangular cartesian coordinate system. Associate to each point of S an arbitrary but fixed unit vector v^i such that $v^i = p^\alpha x^i_{,\alpha}$, $g_{\alpha\beta}p^\alpha p^\beta = 1$, and v^i are real analytic functions of u^α. It will be denoted throughout by v or p at appropriate places. Let $C: u^\alpha = u^\alpha(s)$, $\alpha = 1, 2$ be a curve on S. Then to each point $P(x^i)$ of C there is associated a unit vector v. Consequently, v^i are functions of the arc length s along the curve C. The derived vector of v along C at P is defined by

$$(2.1) \qquad dv^i/ds = {}_vk = {}_v\kappa w^i = x^i_{;\alpha} p^\alpha_{;\gamma} du^\gamma/ds + (p^\delta d_{\delta\gamma} du^\gamma/ds) X^i,$$

* Received January 9, 1952; revised February 4, 1952.

where $p^{\alpha}_{,\gamma}$ is the covariant derivative of the contravariant vector p^{α} based upon $g_{\alpha\beta}$, where X^i are the components of the unit normal to S at P, and where ${}_v\kappa$, called *the absolute curvature of the vector field v with respect to C* at P, is the magnitude of the derived vector ${}_vk$ in whose sense the unit vector w^i is taken. Let ω be the angle between w^i and X^i. We have from (2.1) ${}_v\kappa \cos\omega = p^{\delta} d_{\delta\gamma} du^{\gamma}/ds$. If ${}_v\kappa$ does not vanish, the right side of the above equation is positive or negative according as $0 \leqq \omega < \frac{1}{2}\pi$ or $\pi \geqq \omega > \frac{1}{2}\pi$. Let ${}_v\kappa_n = e p^{\delta} d_{\delta\gamma} du^{\gamma}/ds$, where $e = 1$ is to be taken in the first case, and $e = -1$ in the second. Then ${}_v\kappa_n$ is the magnitude of the normal component in (2.1).

Definition 1. The signed magnitude of the normal curvature vector of a vector field with respect to a curve at P is called *the normal curvature of the vector field with respect to the curve* at P. It will be denoted by $e({}_v\kappa_n)$.

The normal curvature of the field v with respect to C is defined more generally by

$$e({}_v\kappa_n) = d_{\delta\gamma} p^{\delta} du^{\gamma} / (g_{\alpha\gamma} du^{\alpha} du^{\gamma} g_{\lambda\mu} p^{\lambda} p^{\mu})^{\frac{1}{2}} \tag{2.2}$$

when p^{δ} are not necessary the components of a unit vector.

The normal curvatures of v with respect to all curves through P on S are equal if and only if ${}_v\kappa_n$ is independent of du^{α}. Thus $\partial({}_v\kappa_n)/\partial(du^{\alpha}) = 0$ for all du^{α} at P, which can be found to be equivalent to $gd = 0$ at P. Since $g > 0$, and (2.2) is symmetric in p^{α} and du^{α}, we conclude that the normal curvature of any vector field with respect to any curve in the surface is a constant at a point if and only if $d = 0$ at the point. This constant is equal to zero by Definition 3 which follows. Unless otherwise indicated, we shall always assume $d \neq 0$. It is easy to show that there exists a unique direction with respect to which the normal curvature of a vector field at P has a finite extreme value different from zero.

Definition 2. The unique direction, with respect to which the normal curvature of a vector field at P has its extreme value, is known as *the principal direction of the field* at P and the corresponding normal curvature is called *the principal curvature of the field* at P. A curve on S, whose direction at each and every point is a principal direction of a field, is called *a line of curvature of the field.*

The principal direction of the vector field v and the line of curvature of the vector field v at P are found from (2.2) by setting $\partial({}_v\kappa_n)/\partial(du^{\alpha})$ equal to zero. The resulting equation after simplification is

$$\epsilon^{\alpha\beta} g_{\alpha\gamma} d_{\beta\delta} p^{\delta} du^{\gamma} = 0. \tag{2.3}$$

Definition 3. A direction, with respect to which the normal curvature of a vector field is zero, is called an *asymptotic direction of the vector field.* A curve on the surface, whose direction at each and every point is an asymptotic direction of a vector field, is called *an asymptotic line of the vector field.*

The asymptotic direction of the vector field v at P and the asymptotic line of the vector field v at P are defined by the equation

$$d_{\delta\gamma}p^{\delta}du^{\gamma} = 0. \tag{2.4}$$

Definition 4. A curve on the surface, along which the vectors of a vector field of S are tangent to the curve, is called *a curve of the vector field.*

The curve of the vector field v at P is defined by

$$\epsilon_{\alpha\beta}p^{\alpha}du^{\beta} = 0. \tag{2.5}$$

The normal curvature of a vector field with respect to the curve of the field at a point is the normal curvature of the curve at the point. In this respect, the normal curvature of a curve on the surface may be considered as a special case of the normal curvature of a vector field.

Definition 5. A curve on the surface, relative to which the vectors of a given field are parallel in the sense of Levi-Civita, is called *an indicatrix of the vector field.*

The indicatrix of the vector field v at P is defined by

$$p^{\alpha}_{;\gamma}du^{\gamma} = 0. \tag{2.6}$$

Definition 6. A field of vectors, whose directions at each and every point of S constitute a family of principal directions of S, is called *a principal vector field* or *a principal field of S.* The direction of the vector of a principal field and the line of curvature of S at P determined by the principal field are respectively called *the corresponding principal direction of S* and *the corresponding line of curvature of S at P.*

It is obvious that the curve of a principal field at P is the corresponding line of curvature of S at P. The vector field v is a principal field of S if and only if

$$\epsilon^{\alpha\beta}d_{\alpha\gamma}g_{\beta\delta}p^{\gamma}p^{\delta} = 0. \tag{2.7}$$

3. Properties. From Definitions 3 and 4, it is obvious that the asymptotic direction of a vector field is conjugate to the direction of the vector of

the field at P. The orthogonal trajectory of the asymptotic line of the vector field v at P is given by the equation

$$\epsilon_{\alpha\beta} g^{\alpha\gamma} d_{\delta\gamma} p^{\delta} du^{\beta} = 0,$$

which is the same as (2. 3) since $\epsilon_{\alpha\beta} g^{\alpha\gamma} = \epsilon^{\lambda\gamma} g_{\lambda\beta}$. Hence we have

THEOREM 3. 1. *The asymptotic lines of a vector field form a conjugate net with the curves of the field on the surface and form an orthogonal net with the lines of curvature of the field on the surface.*

Let v be a principal vector field. Then on substituting for p^{α} in (2. 7) their expressions from equation (2. 3), we obtain

$$\epsilon^{\beta\gamma} g_{\delta\gamma} d_{\alpha\beta} e^{\alpha\mu} \epsilon^{\lambda\nu} g_{\lambda\sigma} d_{\nu\mu} du^{\sigma} e^{\delta\rho} \epsilon^{\psi\omega} g_{\psi\tau} d_{\omega\rho} du^{\tau} = 0. \tag{3.1}$$

By the application of the process of contraction and the formulae

$$\epsilon^{\alpha\beta} = e^{\alpha\beta}/g^{\frac{1}{2}}, \qquad \epsilon_{\delta\alpha} d^{\delta\beta} = \epsilon^{\sigma\beta} d_{\sigma\alpha}, \qquad d^{\alpha\beta} d_{\beta\gamma} = \delta^{\alpha}{}_{\gamma},$$

$$\epsilon^{\alpha\beta} g_{\alpha\gamma} g_{\beta\delta} = \epsilon_{\gamma\delta}, \qquad \epsilon^{\alpha\gamma} \epsilon^{\beta\delta} d_{\alpha\beta} = d^{\gamma\delta},$$

equation (3. 1) can be reduced to the form $\epsilon^{\alpha\psi} d_{\alpha\sigma} g_{\tau\psi} du^{\sigma} du^{\tau} = 0$, which is the equation of the lines of curvature of S. Since p^{α} can not be orthogonal to du^{α} by Theorem 3. 1 and by the assumption $d \neq 0$, they must be coincident.

THEOREM 3. 2. *The principal direction of a principal vector field of S at P coincides with the corresponding principal direction of S at P. The line of curvature of a principal vector field of S at P is the corresponding line of curvature of S at P. The principal curvature of a principal vector field at P is the corresponding principal curvature of S at P.*

The principal direction of the vector field v at P is found from (2. 3) to be

$$du^{\gamma} = g^{\frac{1}{2}} g^{\gamma\beta} d_{\delta\beta} p^{\delta}. \tag{3.2}$$

The square of the length of this vector is equal to

$$g_{\alpha\beta} du^{\alpha} du^{\beta} = g g^{\beta\lambda} d_{\delta\beta} d_{\mu\lambda} p^{\delta} p^{\mu}. \tag{3.3}$$

Substitution of (3. 2), (3. 3) into (2. 2) gives

$$e({}_{v}\bar{\kappa}_{n}) = (h_{\lambda\delta} p^{\lambda} p^{\delta} / g_{\lambda\delta} p^{\lambda} p^{\delta})^{\frac{1}{2}} \tag{3.4}$$

where $h_{\alpha\beta} = \sum_{i} (\partial X^{i}/\partial u^{\alpha} \cdot \partial X^{i}/\partial u^{\beta})$.

The extreme values of the extreme values (3. 4) at P as the vector field v varies are given by those vector fields v for which $\partial({}_{v}\bar{\kappa}_{n})/\partial p^{\delta} = 0$, that is,

$g_{\alpha\beta}p^\alpha p^\beta h_{\lambda\delta}p^\lambda - h_{\alpha\beta}p^\alpha p^\beta g_{\delta\lambda}p^\lambda = 0$. By means of (3. 4) this equation is expressible in the form

$$(3.5) \qquad (h_{\delta\lambda} - {}_v\bar{\kappa}_n{}^2 g_{\delta\lambda})p^\lambda = 0.$$

Eliminating ${}_v\bar{\kappa}_n{}^2$ from (3. 5) we obtain $\begin{vmatrix} h_{1\lambda}p^\lambda & h_{2\lambda}p^\lambda \\ g_{1\lambda}p^\lambda & g_{2\lambda}p^\lambda \end{vmatrix} = 0$, which, in consequence of [3, p. 253], implies that

$$(3.6) \qquad h_{\alpha\beta} = d_{\alpha\beta}M - g_{\alpha\beta}K, \text{ i. e., } \begin{vmatrix} d_{1\lambda}p^\lambda & d_{2\lambda}p^\lambda \\ g_{1\lambda}p^\lambda & g_{2\lambda}p^\lambda \end{vmatrix} = 0,$$

where M and K are respectively the mean curvature and the Gaussian curvature of S. Hence p^α are the components of the principal directions of S.

THEOREM 3. 3. *The two principal vector fields of S are the vector fields in S which have the extremal principal curvatures of all vector fields in S at P.*

This theorem characterizes a line of curvature of S as a curve whose unit tangent vectors at each and every point form a vector field whose principal curvature assumes an extreme value of the normal curvatures of all vector fields in S.

From Theorems 3. 2 and 3. 3 it is obvious that the extreme values of the principal curvatures of all vector fields in S are equal respectively to the principal curvatures of the surface. Hence

THEOREM 3. 4. *The principal curvatures of the surface at a point are the extremal principal curvatures of all vector fields in the surface at the point.*

Since $h_{\alpha\beta}$ is the fundamental tensor of the Gaussian representation G of the surface S, $(h_{\alpha\beta}p^\alpha p^\beta)^{\frac{1}{2}}$ is the magnitude of the vector p^α in G. If we assume v to be a field of unit vectors in S, then (3. 4) reduces to

$$(3.7) \qquad e({}_v\bar{\kappa}_n) = (h_{\alpha\beta}p^\alpha p^\beta)^{\frac{1}{2}}.$$

Hence we have

THEOREM 3. 5. *The principal curvature of a unit vector field in S at P is numerically equal to the magnitude of the same vector in the Gaussian representation of S at the corresponding point.*

If the vector field does not necessary consist of unit vectors, the content of the above theorem can be stated as follows:

The principal curvature of a vector field in S at P is numerically equal to the ratio of the two magnitudes of the same vector in the Gaussian representation of S and in S at the corresponding points.

If equations (3.6) be multiplied by $p^\alpha p^\beta$ and summed up with respect to α and β, we have

$$(3.8) \qquad ({}_v\bar{\kappa}_n)^2 = M\kappa_n - K,$$

where κ_n is the normal curvature of the curve of the field v on S. Hence

THEOREM 3.6. *The square of the principal curvature of a vector field in S at P is a number diminished by the Gaussian curvature of S at P, the number being the product of the mean curvature of S and the normal curvature of the curve of the field at P.*

Let p^α and q^α be two orthogonal vector fields. Let $e({}_p\bar{\kappa}_n)$ and $e({}_q\bar{\kappa}_n)$ denote respectively the principal curvatures of the two fields. Let κ_1 and κ_2 be the principal curvatures of the surface S. Then by (3.8) and by Euler's theorem we have

$$(3.9) \qquad ({}_p\bar{\kappa}_n)^2 + ({}_q\bar{\kappa}_n)^2 = M^2 - 2K = \kappa_1{}^2 + \kappa_2{}^2.$$

Hence

THEOREM 3.7. *The sum of the squares of the principal curvatures of two orthogonal vector fields at P is constant and is independent of the choice of the two orthogonal vector fields. The constant is equal to the square of the mean curvature of S diminished by twice of the Gaussian curvature of S at P or equal to the sum of the squares of the principal curvatures of S at P.*

Similarly, from (3.8) we find

$$(3.10) \qquad ({}_p\bar{\kappa}_n \cdot {}_q\bar{\kappa}_n)^2 = K^2 + \tfrac{1}{4}M^2(\kappa_1 - \kappa_2)^2 \sin^2 2\theta,$$

where θ is the angle between a line of curvature of S at P and the curve of one of the two fields at P.

By Theorem 3.4 it is evident that if P is an elliptic point, the principal curvatures of all vector fields at P are of the same sign. If P is a hyperbolic point, the principal curvatures of certain vector fields are positive and the principal curvatures of other vector fields are negative.

At a hyperbolic point the normal curvatures of a vector field with respect to different curves are either all positive or all negative, because the normal curvature of a vector field is zero with respect to the asymptotic line of the field, which is orthogonal to the principal direction of the field, and because ${}_v\kappa_n$ is continuous at every point of S. Hence the principal curvature of a vector field at either an elliptic point or a hyperbolic point is sufficient for the determination of the sign of the normal curvatures of the vector field with respect to different curves. As a vector field at the hyperbolic point P

varies from one principal direction of S to another principal direction of S through an asymptotic direction of S, the principal curvature of the field changes from $+$ to $-$ or vice versa. Hence the principal curvature of a vector field has the sign of that principal curvature of S which is situated together with the vector of the field at P in the same section separated by the asymptotic lines of S at P. Consequently, at a hyperbolic point of S the principal curvatures of the two orthogonal vector fields may or may not differ in sign. From (3.10) we have

THEOREM 3.8. *At an elliptic or hyperbolic point P of S, the product of the principal curvatures of two orthogonal vector fields is respectively $\geqq$ or numerically $\geqq$ the Gaussian curvature of S at P. The equality sign holds if and only if P is an umbilic point of S, or the two vector fields are the principal fields of S, or S is a sphere or a minimal surface.*

Let $\lambda_1|^\alpha$ be a unit vector in the asymptotic direction of the vector field v at P and $\lambda_2|^\alpha$ a unit vector in the principal direction of the vector field v at P on S. Then we have $g_{\alpha\beta}\lambda_1|^\alpha\lambda_2|^\beta = 0$, $g_{\alpha\beta}\lambda_\gamma|^\alpha\lambda_\gamma|^\beta = 1$; $\gamma = 1, 2$. Any unit vector tangential to S at P, say du^α, can be expressed as a linear combination of $\lambda_1|^\alpha$ and $\lambda_2|^\alpha$, such as

$$du^\alpha = \lambda_1|^\alpha \cos\theta_1 + \lambda_2|^\alpha \cos\theta_2, \tag{3.11}$$

where $\cos\theta_\gamma = g_{\alpha\beta}\lambda_\gamma|^\alpha du^\beta$; $\gamma = 1, 2$. Assuming $g_{\alpha\beta}p^\alpha p^\beta = 1$ and making use of (2.2) and Definitions 2 and 3, we find the normal curvature of the vector field v with respect to a curve with (3.11) as its tangent vector at P is equal to $e({}_v\kappa_n) - e({}_v\bar{\kappa}_n)\cos\theta_2$. Thus the normal curvature of a vector field with respect to an arbitrary curve can be expressed in terms of the principal curvature of the field. Hence we have the following property similar to Euler's and Meusnier's theorems.

THEOREM 3.9. *The normal curvature of a vector field with respect to an arbitrary curve C always satisfies the following first relation with the principal curvature of the field and the second relation with the absolute curvature of the field with respect to C:*

$${}_v\kappa_n = {}_v\bar{\kappa}\cos\phi, \qquad {}_v\kappa_n = {}_v\kappa\cos\omega,$$

where ϕ is the angle between the direction of the curve C and the principal direction of the field and ω denotes the angle between w^i and X^i at the point considered.

Let a vector field in S be given. Through a point P on S, there pass the following two curves: the curve of the field and the asymptotic line of the field. The straight lines along these two curves in the direction of the vectors of the field associated to the curve form generators of two developable surfaces. The first one is the tangential surface of the curve of the field and the second one is developable because of the conjugacy between the asymptotic direction of the field and the vector of the field. We want to show that these two curves are the only curves possessing this property on S.

Let C: $u^\alpha = u^\alpha(s)$ be a curve on S such that the straight lines on v associated with C are generators of a developable surface. Then the developable surface will be defined by

$$y^i = x^i(s) + \lambda v^i, \tag{3.12}$$

where s and λ are parameters. $\lambda = a$ constant $\neq 0$ defines a curve on the developable surface, the tangent of which has the direction defined by $\partial y^i/\partial s = t^i + \lambda \cdot {}_v k$, where ${}_v k$ has the same meaning as in (2.1).

Since $(t^i, v^i, t + \lambda \cdot {}_v k) = 0$ implies $(t^i, v^i, {}_v k) = 0$, we find that a necessary and sufficient condition for the surface (3.12) to be a developable is ${}_v k = 0$ or $t^i = \pm v^i$.

Case 1. ${}_v k = 0$. Then $v^i =$ const. The vectors of the field are absolutely parallel. The lines on them form generators of a developable surface along any curve on the surface. The surface is consequently isometric with the plane. If $v^i \neq$ const., then ${}_v\kappa = 0$ and from (2.1) we have $p^\alpha_{;\gamma} du^\gamma = 0$, $p^\delta d_{\delta\gamma} du^\gamma = 0$, which mean geometrically that the lines on the vectors of the field will form generators of a developable surface along the curve C which is both an indicatrix of the field and an asymptotic line of the field.

Case 2. $t^i = \pm v^i$. The curve C is then a curve of the vector field. Hence

THEOREM 3.10. *At a point on a non-developable surface there are two and only two curves, along each of which the straight lines on the vectors of a field form generators of a developable surface. These curves are the asymptotic line of the field and the curve of the field.*

The following theorem is an immediate consequence of Definitions 1 and 5:

THEOREM 3. 11. *The normal curvature of a vector field at P is the same with respect to all curves tangent at P to the same direction. The normal curvature of either one of two vector fields with respect to the curve of the other field is the same. The absolute curvature vector of a vector field with respect to the indicatrix of the field has at each point the direction of the surface normal. The absolute curvature and the normal curvature of a vector field with respect to a curve C are numerically equal if and only if the curve C is the indicatrix of the field.*

4. Generalization. Let $V_n: y^\alpha = y^\alpha(x^1, \cdots, x^n)$, $\alpha = 1, \cdots, n+1$ be a hypersurface in V_{n+1}. Let $C: x^i = x^i(s)$, $i = 1, \cdots, n$ be a curve on V_n. Let v be a field of unit vectors in V_n such that $v^\alpha = p^i y^\alpha_{,i}$ and v^α are real analytic functions of x^i. Then *the normal curvature of v with respect to C at a point P of V_n* is defined by

$$e({}_v\kappa_\xi) = \Omega_{ij} p^i dx^j / (g_{ij} dx^i dx^j g_{hl} p^h p^l)^{1/2}. \tag{4. 1}$$

We assume the fundamental forms of all spaces discussed in this section are positive definite.

When, in particular, the vector v is tangent to C at P,

$$e({}_v\kappa_\xi) = \Omega_{ij} dx^i dx^j / g_{ij} dx^i dx^j$$

and, consequently, the normal curvature of v with respect to C becomes the normal curvature of the curve C. It is evident ${}_v\kappa_\xi = 0$ at P along the direction for which

$$\Omega_{ij} p^i dx^j = 0 \tag{4. 2}$$

This direction is called *an asymptotic direction of the vector field v at P of V_n*. A curve, whose direction at each and every point is an asymptotic direction of a vector field is *an asymptotic line of the field.* Such a direction and such a line satisfy the differential equation (4. 2).

The extreme values of $e({}_v\kappa_\xi)$ at P are attained with respect to those directions for which $\partial({}_v\kappa_\xi)/\partial(dx^i) = 0$ for $i = 1, \cdots, n$, that is

$$(\Psi_{ij} - {}_v\kappa_\xi^2 g_{ij}) dx^i = 0, \tag{4. 3}$$

where

$$\Psi_{ij} = \Omega_{ij} \Omega_{hl} p^h p^l / g_{ij} p^i p^j \tag{4. 4}$$

and where ${}_v\kappa_\xi^2$ is a root of the equation

$$| \Psi_{ij} - {}_v\kappa_\xi^2 g_{ij} | = 0. \tag{4. 5}$$

Since the metric of V_n is assumed to be positive definite, the roots of (4. 5) are all real; and since $\| \Psi_{ij} \|$ is of rank 1, $n-1$ roots are equal to zero which correspond to asymptotic directions of the vector field v at P. The square of the non-zero extreme value corresponds by (4. 3) to a principal direction determined by the tensor Ψ_{ij}. The extreme value of $e({}_v\kappa_\xi)$ is called *the principal normal curvature of the vector field v at P in V_n* and the corresponding direction is *the principal direction of the field at P in V_n.*

From (4. 1) we know that *the normal curvature of a vector field in V_n has equal value with respect to all curves of V_n through P with the same direction. Its value with respect to an indicatrix of the field is the absolute curvature of the field in V_{n+1} with respect to the indicatrix of the field of V_n,* since *an indicatrix of the field v* satisfies $p^i_{,j}dx^j = 0$, and ${}_v\kappa$, the *absolute curvature of the field v with respect to C,* is defined by

$$ {}_vk = {}_v\kappa w^\alpha = dv^\alpha/ds = p^i_{,j} y^\alpha_{,i} dx^j/ds + e({}_v\kappa_\xi)\xi^\alpha, $$

where w^α is a unit vector in the same sense as ${}_vk$. Similarly, *when an asymptotic line* (4. 2) *of a vector field v is an indicatrix of the field in V_n, the latter is also the indicatrix of the field in the enveloping space V_{n+1} and conversely.*

The principal direction of the vector field v in V_n is indeterminate if and only if

$$ (4.\ 6) \qquad \Psi_{ij} = {}_v\kappa_\xi^2 g_{ij}. $$

Such space is then homogeneous with respect to the tensor Ψ_{ij}. *An asymptotic line of a vector field is conjugate to a curve of the field at P in V_n.*

If the vector field v is a principal vector field in V_n, then

$$ (4.\ 7) \qquad (\Omega_{ij} - \kappa_\xi g_{ij}) p^i = 0, $$

where κ_ξ is the corresponding principal curvature of V_n at P. In virtue of (4. 1), (4. 3), (4. 4) and 4. 7) we obtain $e({}_v\kappa_\xi) = \kappa_\xi$ at P. It follows that:

The principal direction of a principal vector field in V_n at P and the principal curvature of the same field in V_n at P are respectively identical with the corresponding principal direction of V_n and the corresponding principal curvature of V_n at P. Consequently, *a line of curvature of a principal vector field in V_n is the corresponding line of curvature of V_n.*

Let the unit vectors of an orthogonal ennuple at P formed by $\lambda_h|^i$ be taken in the principal direction of the vector field v at P in V_n and

asymptotic directions of the vector field v at P in V_n. In particular, let $\lambda_1|^i$ be in the principal direction. Then we have

$$(4.8)\qquad \begin{aligned} &g_{ij}\lambda_h|^i\lambda_h|^j = 1, && g_{ij}\lambda_h|^i\lambda_k|^j = 0, \quad h \neq k \\ &\Psi_{ij}\lambda_h|^i\lambda_h|^j = {}_v\kappa_\xi^2, && \Psi_{ij}\lambda_h|^i\lambda_k|^j = 0, \quad h \neq k. \end{aligned}$$

Any field of unit vectors in V_n, say dx^i, is defined by

$$(4.9)\qquad dx^i = \lambda_1|^i \cos\theta_1 + \cdots + \lambda_n|^i \cos\theta_n,$$

where $\cos\theta_r = g_{ij}\lambda_r|^i dx^j$. Thinking that p^i and dx^i are unit vectors, we can write (4.1) in the form $e({}_v\kappa_\xi) = \Omega_{ij}p^i dx^j$, and by (4.4) we have

$$(4.10)\qquad ({}_v\kappa_\xi)^2 = \Psi_{ij}dx^i dx^j.$$

Substituting in (4.10) from (4.9) and making use of (4.3) and (4.8), *we obtain*

$$(4.11)\qquad {}_v\kappa_\xi^2 = {}_v\bar{\kappa}_\xi^2 \cos^2\theta_1$$

which is analogous to Euler's formula, $e({}_v\bar{\kappa}_\xi)$ being the principal curvature of the vector field.

Consider a Riemannian space V_n of coordinates x^i immersed in a Riemannian space V_m of coordinates y^α. Then *the normal curvature of the vector field v at P in V_n in V_m* is denoted by $e({}_v\kappa_\xi)$ such that

$$(4.12)\qquad ({}_v\kappa_\xi)^2 = (\sum_\nu \Omega_\nu|_{ij}\Omega_\nu|_{hl}p^i p^h dx^j dx^l)/(g_{ij}g_{hl}p^i p^j dx^h dx^l),$$

where $i, j, h, l = 1, \cdots, n$, $\nu = n+1, \cdots, m$. Proceeding with this definition and others in a manner similar to the preceding discussion for V_n in V_{n+1}, we may obtain for V_n in V_m the following properties:

The normal curvature at a point P of a vector field with respect to a curve C in V_n in V_m depends only on the direction dx^i of C at P, and is the same with respect to all curves tangent to C at P in V_n. It is zero in an asymptotic direction of the field, which is conjugate to the vector of the field, and its finite extreme value for the normal $\xi_\nu|^\alpha$ is attained with respect to the principal direction of the field for the normal $\xi_\nu|^\alpha$. The principal direction at P of a principal vector field in V_n and the principal normal curvature at P of the same field for the normal $\xi_\nu|^\alpha$ are respectively identical with the corresponding principal direction of V_n at P and the corresponding curvature of V_n at P for the normal $\xi_\nu|^\alpha$. A line of curvature of a principal vector field in V_n for the normal $\xi_\nu|^\alpha$ is the line of curvature determined by this

corresponding family of principal directions of V_n for the normal $\xi_r|^\alpha$. The normal curvature of a vector field with respect to an indicatrix of the field in V_n. An asymptotic line of a vector field in V_n and an indicatrix of the field in V_n are coincident if and only if the latter is also an indicatrix of the field in V_m.

UNIVERSITY OF CALIFORNIA, BERKELEY.

BIBLIOGRAPHY.

[1] W. C. Graustein, "Parallelism and equidistance in classical differential geometry," *Transactions of the American Mathematical Society*, vol. 34 (1932), pp. 557-593.

[2] R. M. Peters, "Parallelism and equidistance in Riemannian geometry," *American Journal of Mathematics*, vol. 5 (1935), pp. 103-111.

[3] L. P. Eisenhart, *An introduction to differential geometry with the use of the tensor calculus*, Princeton University Press, 1947.

[4] L. P. Eisenhart, *Riemannian geometry*, Princeton University Press, 1949.

A SYNTHESIS OF THE CLIFFORD MATRICES AND ITS GENERALIZATION.* [1]

By T. G. Room.

Clifford matrices of 2^r row and columns are a set of $2r+1$ matrices γ_i, with the properties

$$\gamma_i^2 = I, \qquad \gamma_i\gamma_j + \gamma_j\gamma_i = 0.$$

In Part I these matrices are synthesized from two elementary sets of r matrices, U_i which are diagonal matrices, and V_i which are substitution matrices. The key to the development is the numbering of the rows and columns of the matrix in the reversed binary scale, namely as $\alpha_1 + 2\alpha_2 + \cdots + 2^{r-1}\alpha_r$.

In Part II, geometrical properties of the collineations determined by these matrices are worked out, and in particular the existence of a sequence of systems of spaces exhibiting a "Mutuality" exactly analogous to Study's "Triality" is established.

Part III suggests generalizations of Part I using scales other than the binary to obtain sets of matrices in which most pairs commute, but for some pairs, in one case $UV = kVU$, and in another $U^2V = VU$.

Part I.

1. The basic family of matrices. Number the rows and columns of a $2^r \times 2^r$ matrix in the reversed binary scale, beginning with 0, so that the row or column in position number $\alpha_1 + 2\alpha_2 + 4\alpha_3 + \cdots + 2^{r-1}\alpha_r$ from the top or the left carries the indices $\alpha_1\alpha_2\alpha_3 \cdots \alpha_r$. Throughout Parts I and II of this paper I use the conventions

(i) Greek indices take the values 0 and 1 only,

(ii) All arithmetical functions of Greek indices are reduced modulo 2,

(iii) For any index α_i, α'_i is defined by $\alpha_i + \alpha'_i = 1$.

* Received November 22, 1950; revised July 25, 1952.

[1] My interest in the Clifford matrices was aroused by Professor O. Veblen, and Drs. W. J. Givens and N. H. Kuiper. A recently published paper by Dr. Kuiper (1949) gives much of the geometrical background, and many references to the history of the problem may be found in it. Part II of this paper was presented to the International Congress of Mathematicians.

Denote the elements of a matrix by $a^{\alpha_1\alpha_2\cdots\alpha_r}_{\beta_1\beta_2\cdots\beta_r}$ or $a^{(\alpha_i)}_{(\beta_j)}$ (with superscript as row number) and define matrices $\boldsymbol{U}_i$, $\boldsymbol{V}_i$ thus:

$\boldsymbol{U}_i$ is the diagonal matrix with elements

$$a^{\alpha_1\alpha_2\cdots\alpha_r}_{\alpha_1\alpha_2\cdots\alpha_r} = (-1)^{\alpha_i}$$

$\boldsymbol{V}_i$ is the substitution matrix with only one non-zero element in each row and column, given by

$$a^{\alpha_1\alpha_2\cdots\alpha_{i-1}\alpha_i\alpha_{i+1}\cdots\alpha_r}_{\alpha_1\alpha_2\cdots\alpha_{i-1}\alpha'_i\alpha_{i+1}\cdots\alpha_r} = 1,$$

where each superscript except the i-th is equal to the corresponding subscript, and the i-th superscript and subscript are different.

For $r = 2$ the matrices are

$$\boldsymbol{U}_1 = \begin{bmatrix} 1 & & & \\ & -1 & & \\ & & 1 & \\ & & & -1 \end{bmatrix}, \qquad \boldsymbol{U}_2 = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & -1 & \\ & & & -1 \end{bmatrix},$$

$$\boldsymbol{V}_1 = \begin{bmatrix} & 1 & & \\ 1 & & & \\ & & & 1 \\ & & 1 & \end{bmatrix}, \qquad \boldsymbol{V}_2 = \begin{bmatrix} & & 1 & \\ & & & 1 \\ 1 & & & \\ & 1 & & \end{bmatrix}.$$

These matrices have the properties:

THEOREM 1. 1. (i) $\boldsymbol{U}_i^2 = \boldsymbol{V}_i^2 = I$,

(ii) *All pairs commute, except the r pairs $\boldsymbol{U}_i$, $\boldsymbol{V}_i$,*

(iii) $\boldsymbol{U}_i\boldsymbol{V}_i = -\boldsymbol{V}_i\boldsymbol{U}_i$.

THEOREM 1. 2. $\boldsymbol{W}^{(\alpha_i)}_{(\beta_j)}$, $= \Pi\boldsymbol{U}_i^{\alpha_i}\Pi\boldsymbol{V}_j^{\beta_j}$, *is the matrix with one non-zero element in each row and column, given by*

$$a^{(\lambda_i)}_{(\lambda_i+\beta_j)} = (-1)^{\Sigma\alpha_i\lambda_i}.$$

E. g., when $r = 2$,

$$\boldsymbol{W}^{11}_{01} = \boldsymbol{U}_1\boldsymbol{U}_2\boldsymbol{V}_2 = \begin{bmatrix} & & 1 & \\ & & & -1 \\ -1 & & & \\ & 1 & & \end{bmatrix}.$$

THEOREM 1. 3. $$\boldsymbol{W}^{(\alpha_i)}_{(\beta_j)}\boldsymbol{W}^{(\gamma_i)}_{(\delta_j)} = (-1)^{\Sigma\beta_j\gamma_j}\boldsymbol{W}^{(\alpha_i+\gamma_i)}_{(\beta_j+\delta_j)}.$$

THEOREM 1. 4. $$[W_{(\beta_i)}^{(\alpha_i)}]^2 = (-1)^{\Sigma\alpha_i\beta_i} I.$$

If we take $W_{(\beta_k)}^{(\alpha_j)*} = i^{\Sigma\alpha_j\beta_j} W_{(\beta_k)}^{(\alpha_j)}$, $\Sigma\alpha_j\beta_j$ being reduced modulo 2, then

$$[W_{(\beta_k)}^{(\alpha_j)*}]^2 = I.$$

THEOREM 1. 5. *Two matrices* $W_{(\beta_j)}^{(\alpha_i)}$, $W_{(\delta_j)}^{(\gamma_i)}$ *commute or anti-commute according as* $\Sigma\alpha_i\delta_i + \Sigma\beta_j\gamma_j = 0$, *or* 1.

THEOREM 1. 6. $W_{(\beta_i)}^{(\alpha_i)}$ *is symmetric or skew-symmetric according as* $\Sigma\alpha_i\beta_i = 0$, *or* 1.

Thus, the non-zero elements of the matrix are

$$a_{(\lambda_j+\beta_j)}^{(\lambda_i)} = (-1)^{\Sigma\alpha_i\lambda_j}.$$

Since the element $a_{(\lambda_i+\beta_i)}^{(\lambda_i)} = a_{(\lambda_i+\beta_i+\beta_i)}^{(\lambda_i+\beta_i)}$, it is also a non-zero element of the matrix, and

$$a_{(\lambda_i)}^{(\lambda_i+\beta_i)} = (-1)^{\Sigma\alpha_i(\lambda_i+\beta_i)} = (-1)^{\Sigma\alpha_i\beta_i} a_{(\lambda_i+\beta_i)}^{(\lambda_i)}.$$

THEOREM 1. 7. *Any given* $2^r \times 2^r$ *matrix*, $H = (h_{(\beta_j)}^{(\alpha_i)})$, *may be expressed in a single way as* $\Sigma k_{(\beta_j)}^{(\alpha_i)} W_{(\beta_j)}^{(\alpha_i)}$, *there being* 2^{2r} *terms in the summation.*

Thus, if we rearrange the terms in the rows of the matrix H, so that they form the matrix $\tilde{H} = (h_{(\alpha_j+\beta_j)}^{(\alpha_i)})$, then $\tilde{H} = \Omega K$ where $K = (k_{(\beta_i)}^{(\alpha_i)})$ and $\Omega = \Sigma W_{\alpha'_1\alpha'_2\ldots\alpha'_r}^{\alpha_1\alpha_2\ldots\alpha_r}$, that is, Ω is the matrix each of whose elements is $+1$ or -1, the sign being given by $a_{\beta_1\ldots\beta_r}^{\alpha_1\ldots\alpha_r} = (-1)^{\Sigma\alpha_i\beta_i}$.

For $r = 2$, we have

$$\begin{bmatrix} h_{00}^{00} & h_{10}^{00} & h_{01}^{00} & h_{11}^{00} \\ h_{10}^{10} & h_{00}^{10} & h_{11}^{10} & h_{01}^{10} \\ h_{01}^{01} & h_{11}^{01} & h_{00}^{01} & h_{10}^{01} \\ h_{11}^{11} & h_{01}^{11} & h_{10}^{11} & h_{00}^{11} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} k_{00}^{00} & k_{10}^{00} & k_{01}^{00} & k_{11}^{00} \\ k_{00}^{10} & k_{10}^{10} & k_{01}^{10} & k_{11}^{10} \\ k_{00}^{01} & k_{10}^{01} & k_{01}^{01} & k_{11}^{01} \\ k_{00}^{11} & k_{10}^{11} & k_{01}^{11} & k_{11}^{11} \end{bmatrix}$$

$$\Omega = U_1U_2 + V_1U_2 + U_1V_2 + V_1V_2.$$

2. The Clifford matrices. The Clifford matrices $\gamma_i^{(r)}$, of 2^r rows and columns, have the properties:

$$\gamma_i^2 = I, \qquad \gamma_i\gamma_j + \gamma_j\gamma_i = O.$$

They may be defined inductively from the matrix $\gamma_0{}^{(0)}$ of one element thus:

$$\gamma_0{}^{(0)} = (1)\,; \quad \gamma_j{}^{(r)} = \begin{pmatrix} \gamma_j{}^{(r-1)} & \\ & -\gamma_j{}^{(r-1)} \end{pmatrix},$$

$$\gamma_{2r-1}{}^{(r)} = i\begin{pmatrix} & I \\ -I & \end{pmatrix}, \quad \gamma_{2r}{}^{(r)} = \begin{pmatrix} & I \\ I & \end{pmatrix}.$$

In the notation used in this paper, this set is

$$\gamma_0{}^{(r)} = W^{11\ldots1}_{00\ldots0}, \qquad \gamma_1{}^{(r)} = iW^{11\ldots1}_{100..0}, \qquad \gamma_2{}^{(r)} = W^{011..1}_{100..0},$$

$$\gamma_3{}^{(r)} = iW^{011\ldots1}_{010\ldots0}, \qquad \gamma_4{}^{(r)} = W^{0011\ldots1}_{0100\ldots0}, \cdots,$$

$$\gamma^{(r)}_{2r-1} = iW^{000\ldots01}_{000\ldots01}, \qquad \gamma^{(r)}_{2r} = W^{00\ldots00}_{00\ldots01}\,;$$

that is: $\gamma^{(r)}_{2k} = W^{(\alpha_i)}_{(\beta_j)}$, where $\alpha_1, \cdots, \alpha_k = 0$, $\alpha_{k+1}, \cdots, \alpha_r = 1$, and all $\beta_j = 0$, except $\beta_k = 1$, and $\gamma^{(r)}_{2k+1} = iW^{(\alpha_i)}_{(\beta_j)}$ where $\alpha_1, \cdots, \alpha_k = 0$, $\alpha_{k+1}, \cdots, \alpha_r = 1$, and all $\beta_j = 0$, except $\beta_{k+1} = 1$.

For our imemdiate purpose the factor i is unimportant, and we define as the *primary Clifford set*, the set of $2r+1$ matrices W which correspond to the matrices $\gamma^{(r)}$. A *Clifford set* of matrices W is one with the same properties as the $\gamma^{(r)}$, namely $W^2 = \pm I$, $W_jW_k = -W_kW_j$, that is, a set of matrices W such that no pair commutes.

One of the problems (the solution of which I do not attempt here) is the determination of all Clifford sets. It is clear that from any given set we may obtain other sets by any of the operations

(i) changing each U into a V and each V into a U, i. e., by replacing each $W^{(\alpha_i)}_{(\beta_j)}$ by $W^{(\beta_j)}_{(\alpha_i)}$. E. g., the second set of those listed below is obtained in this way from the first,

(ii) effecting on each symbol of the given set the same permutation of both superscripts and subscripts. E. g., the third set below is obtained from the first by interchanging the first and second indices,

(iii) adjoining I to the set, multiplying by any member of the set, and removing I from the resulting set. E. g., the second set below is obtained from the first by multiplying by W^{01}_{01}.

From Theorem 1.5 it follows that the problem is equivalent to the following combinatorial problem:

Let $(\alpha_i{}^{(s)}; \beta_j{}^{(s)})$, $s = 1, \cdots, m$, be m pairs of permutations of r terms

each 0 or 1. It is required to determine all sets of m such pairs of permutations, which are such that for every two pairs in a set

$$\sum_i \alpha_i^{(s)}\beta_i^{(t)} + \sum_i \alpha_i^{(t)}\beta_i^{(s)} \text{ is odd.}$$

In the solutions above $m = 2r + 1$, and this is probably the largest value of m.

In the case $r = 2$, there are six Clifford sets, namely

$$\begin{array}{ccccc}
W^{11}_{00} & W^{11}_{10} & W^{01}_{10} & W^{01}_{01} & W^{00}_{01} \\
W^{00}_{11} & W^{10}_{11} & W^{10}_{01} & W^{01}_{01} & W^{01}_{00} \\
W^{11}_{00} & W^{11}_{01} & W^{10}_{01} & W^{10}_{10} & W^{00}_{10} \\
W^{00}_{11} & W^{01}_{11} & W^{01}_{10} & W^{10}_{10} & W^{10}_{00} \\
W^{10}_{00} & W^{00}_{10} & W^{11}_{11} & W^{11}_{10} & W^{10}_{11} \\
W^{01}_{00} & W^{00}_{01} & W^{11}_{11} & W^{11}_{01} & W^{01}_{11} .
\end{array}$$

If these matrices are regarded as collineation matrices in projective [3], then: the five points obtained from any general point under the transformations of one of the sets lie in a plane and on a conic with the point, and the six such planes are the planes through a point of a Kummer 16_6 configuration.

Part II.

Quadrics associated with Clifford Matrices.

3. The Clifford involutions. For convenience of exposition, in this Part of the work I take $r = 4$.

As coordinates in [15] take $(x) = x_{\alpha\beta\gamma\delta}$,[2] where $\alpha, \cdots, \delta = 0, 1$, and $\alpha\beta\gamma\delta$ are the numbers 0 to 15 expressed as $\alpha + 2\beta + 4\gamma + 8\delta$. The primary Clifford matrices are

$$W^{1111}_{0000};\ W^{1111}_{1000},\ W^{0111}_{1000};\ W^{0111}_{0100},\ W^{0011}_{0100};\cdots;\ W^{0001}_{0001},\ W^{0000}_{0001}.$$

The pairs with identical subscripts have the non-zero elements, ± 1, in the same places, with half the signs agreeing. It is simpler therefore to replace these matrices by the set of 9:

[2] The coordinates $x_{\alpha\beta\gamma\delta}$ are related to Cartan's (Cartan, 1938, p. 4) coordinates $\xi_{ijk\ldots}$ thus: $i, j, k, \cdots$ are the position numbers in any order of the non-zero subscripts $\alpha, \beta, \gamma, \delta$. E. g., x_{1011} is the same as ξ_{134}.

$$Y = W^{1111}_{0000},\ Y^1_0 = \tfrac{1}{2}(W^{1111}_{1000} + W^{0111}_{1000}),$$
$$Y^1_1 = \tfrac{1}{2}(-W^{1111}_{1000} + W^{0111}_{1000}) = (Y^1_0)^T,$$

$$Y^2_0 = \tfrac{1}{2}(W^{0111}_{0100} + W^{0011}_{0100}), \cdots, Y^4_1 = \tfrac{1}{2}(-W^{0001}_{0001} + W^{0000}_{0001}).$$

Each of these has only 8 non-zero elements, which for Y_0 are above the leading diagonal and for Y_1 are below it. The non-zero elements of typical matrices are:

$$\begin{aligned}
Y^1_0:&\quad a^{0\beta\gamma\delta}_{1\beta\gamma\delta} = (-1)^{\beta+\gamma+\delta}\\
Y^1_1:&\quad a^{1\beta\gamma\delta}_{0\beta\gamma\delta} = (-1)^{\beta+\gamma+\delta}\\
Y^2_0:&\quad a^{\alpha 0\gamma\delta}_{\alpha 1\gamma\delta} = (-1)^{\gamma+\delta}\\
Y^3_0:&\quad a^{\alpha\beta 0\delta}_{\alpha\beta 1\delta} = (-1)^{\delta}\\
Y^4_0:&\quad a^{\alpha\beta\gamma 0}_{\alpha\beta\gamma 1} = 1.
\end{aligned}$$

The nine matrices determine a linear family of matrices $\mathbf{\Gamma}$ of freedom 8 defined by

$$\mathbf{\Gamma} = cY + \Sigma c^i_0 Y^i_0 + \Sigma c^i_1 Y^i_1.$$

From this it is easy to prove that

THEOREM 3.1. *$(\mathbf{\Gamma})^2 = \mathcal{B}I$, where $\mathcal{B} = (c)^2 + \Sigma c^i_0\, c^i_1$.*

THEOREM 3.2. *If $\mathcal{B} = 0$, $\mathbf{\Gamma}$ is of rank* 8.

THEOREM 3.3. $\mathbf{\Gamma}^T = cY + \Sigma c^i_1\, Y^i_0 + \Sigma c^i_0\, Y^i_1$.

The matrix $\mathbf{\Delta}$, $= W^{1010}_{1111}$,[8] which is the product of the four skew-symmetric matrices of the original set of 9 Clifford matrices, has the following relations to each $\mathbf{\Gamma}$:

THEOREM 3.4. $\qquad \mathbf{\Delta\Gamma\Delta} = \mathbf{\Gamma}^T.$

THEOREM 3.5. $\qquad \mathbf{\Gamma}^T\mathbf{\Delta\Gamma} = \mathcal{B}\mathbf{\Delta}.$

The first of these comes directly from relations such as

$$\mathbf{\Delta} W^{1111}_{1000}\mathbf{\Delta} = -W^{1111}_{1000},$$

[8] W^{1010}_{1111} is the reversed diagonal matrix with elements $a^{\alpha\beta\gamma\delta}_{\alpha'\beta'\gamma'\delta'} = (-1)^{\alpha+\gamma}$. The corresponding matrix associated with the set of $2r+1$ Clifford matrices of 2^r rows and columns is $W^{101010\ldots}_{111111\ldots}$, and is symmetrical or skew-symmetrical according as the number of non-zero superscripts is even or odd. I. e., the corresponding correlation is a polarity with regard to a quadric if $r = 4s-1$, $4s$, and with regard to a null-system if $r = 4s+1$, $4s+2$.

using the form above, and for the second we have

$$\mathbf{\Gamma}^T\mathbf{\Delta}\mathbf{\Gamma} = \mathbf{\Gamma}^T(\mathbf{\Gamma}^T\mathbf{\Delta}) = (\mathbf{\Gamma}^T)^2\mathbf{\Delta} = \mathscr{G}\mathbf{\Delta}.$$

If $\mathbf{\Gamma}$ is taken to be the matrix of a collineation, it follows from Theorem 3.1, that (i) the collineation is involutory, and (ii) if $\mathscr{G} = 0$, then the collineation is singular, and the invariant space is a $[7]$, γ, say. From Theorem 3.5 it follows that the conjugate, $\mathbf{\Gamma}\boldsymbol{x}$, of a point $\boldsymbol{x}$, is such that $(\boldsymbol{x}^T\mathbf{\Gamma}^T)\mathbf{\Delta}\mathbf{\Gamma}\boldsymbol{x} = \mathscr{G}\boldsymbol{x}^T\mathbf{\Delta}\boldsymbol{x}$. That is

THEOREM 3.6. *The quadric* $Q \equiv \boldsymbol{x}^T\mathbf{\Delta}\boldsymbol{x} = 0$ *is invariant under all the collineations* $\mathbf{\Gamma}$.

THEOREM 3.7. *The invariant space,* γ, *of a singular collineation* $\mathbf{\Gamma}$ *lies on* Q.

4. The space of conjugates of a point. The conjugates of a point P, $(p_{\alpha\beta\gamma\delta})$, under the linear family of Clifford collineations $\mathbf{\Gamma}$ are linearly dependent on the nine points $\bar{P} = \mathbf{Y}P$, $P_\lambda^i = \mathbf{Y}_\lambda^i P$, where, e. g., P_1^2 is given by $x_{\alpha 0\beta\gamma} = 0$, $x_{\alpha 1\beta\gamma} = (-1)^{\beta+\gamma}p_{\alpha 0\beta\gamma}$. In general these points determine a $[8]$.

Take now the eight primes ϖ_λ^i of which the equations, each involving eight terms, are

$$\varpi_\lambda^1: \quad \Sigma(-1)^{\alpha+\gamma}p_{\lambda\alpha\beta\gamma}x_{\lambda\alpha'\beta'\gamma'} = 0,$$

etc., the fixed index for ϖ_λ^i being in the i-th place, and the sign being determined in each case by $(-1)^{\alpha+\gamma}$. Using the matrices $\boldsymbol{W}$ we may write these equations as

$$\begin{aligned}
\varpi_0^1&: \quad \boldsymbol{p}^T(\boldsymbol{W}_{0111}^{0101} + \boldsymbol{W}_{0111}^{1101})\boldsymbol{x} = 0,\\
\varpi_1^1&: \quad \boldsymbol{p}^T(\boldsymbol{W}_{0111}^{0101} - \boldsymbol{W}_{0111}^{1101})\boldsymbol{x} = 0,\\
\varpi_0^2&: \quad \boldsymbol{p}^T(\boldsymbol{W}_{1011}^{1001} + \boldsymbol{W}_{1011}^{1101})\boldsymbol{x} = 0, \text{ etc.}
\end{aligned}$$

From the following relations we see that ϖ_0^1 contains $\bar{P}$, P_1^1, and the six points P_λ^i, $j = 2, 3, 4$, which may be typified by P_λ^2:

$$\begin{aligned}
\bar{P}&: \quad \boldsymbol{p}^T(\boldsymbol{W}_{0111}^{0101} + \boldsymbol{W}_{0111}^{1101})\boldsymbol{W}_{0000}^{1111}\boldsymbol{p} = -\boldsymbol{p}^T(\boldsymbol{W}_{0111}^{1010} + \boldsymbol{W}_{0111}^{0010})\boldsymbol{p}\\
&\qquad = 0 \text{ since the two matrices } \boldsymbol{W} \text{ are skew-symmetric.}
\end{aligned}$$

$$\begin{aligned}
P_1^1&: \quad \boldsymbol{p}^T(\boldsymbol{W}_{0111}^{0101} + \boldsymbol{W}_{0111}^{1101})(\boldsymbol{W}_{1000}^{1111} - \boldsymbol{W}_{1000}^{0111})\boldsymbol{p}\\
&\qquad = \boldsymbol{p}^T(-\boldsymbol{W}_{1111}^{1010} + \boldsymbol{W}_{1111}^{0010} - \boldsymbol{W}_{1111}^{0010} + \boldsymbol{W}_{1111}^{1010})\boldsymbol{p} = 0
\end{aligned}$$

$$P_\lambda^2:\quad \boldsymbol{p}^T(\boldsymbol{W}_{0111}^{0101}+\boldsymbol{W}_{0111}^{1101})(\boldsymbol{W}_{0100}^{0111}+\boldsymbol{W}_{0100}^{0011})\boldsymbol{p}$$
$$=\boldsymbol{p}^T(-\boldsymbol{W}_{0011}^{0010}+\boldsymbol{W}_{0011}^{0110}-\boldsymbol{W}_{0011}^{1010}+\boldsymbol{W}_{0011}^{1110})\boldsymbol{p}$$
$= 0$ since each of the matrices $\boldsymbol{W}$ is skew-symmetric.

ϖ_0^1 contains P_0^1 if

$$0=\boldsymbol{p}^T(\boldsymbol{W}_{0111}^{0101}+\boldsymbol{W}_{0111}^{1101})(\boldsymbol{W}_{1000}^{1111}+\boldsymbol{W}_{1000}^{0111})\boldsymbol{p}=-2\boldsymbol{p}^T\boldsymbol{\Delta}\boldsymbol{p}.$$

From the corresponding relations among the points $\bar{P}$, P_λ^i, and the primes ϖ_μ^j, it follows that all the conjugates of P lie in all eight primes, if P lies on $Q\equiv \boldsymbol{x}^T\boldsymbol{\Delta}\boldsymbol{x}=0$. I. e.,

THEOREM 4.1. *The necessary and sufficient condition that the space of conjugates of a point P should be of dimension less than 8 is that P should lie on the quadric Q.*

There are two systems of [7]-generators on Q, each of freedom 28, and such that a general pair of spaces of opposite systems has a common point. The invariant [7]'s γ, of the singular collineations belong to one system and have freedom 7, the [7]'s above belong to the other system and have freedom 14.

Take next the 24 primes $\varpi_{\lambda\mu}^{ij}$, typified by

$$\varpi_{\lambda\mu}^{24}:\quad \Sigma(-1)^\alpha p_{\alpha'\lambda\beta'\mu}x_{\alpha\lambda\beta\mu}=0,[4]$$

the sum extending over the four terms given by possible pairs of $\alpha\beta$. These primes contain all nine points $\bar{P}$, P_λ^i, as well as P itself, if, in addition to lying on Q, the point P lies on the 8 quadric [7]-cones:

$$q_\lambda^1\equiv\Sigma(-1)^\alpha x_{\lambda\alpha\beta0}x_{\lambda\alpha'\beta'1}=0,$$
$$q_\lambda^2\equiv\Sigma(-1)^\alpha x_{\alpha\lambda\beta0}x_{\alpha'\lambda\beta'1}=0,$$
$$q_\lambda^3\equiv\Sigma(-1)^\alpha x_{\alpha\beta\lambda0}x_{\alpha'\beta'\lambda1}=0,$$
$$q_\lambda^4\equiv\Sigma(-1)^\alpha x_{\alpha\beta0\lambda}x_{\alpha'\beta'1\lambda}=0.[5]$$

In terms of the matrices $\boldsymbol{W}$ these quadrics are determined by:

$$q_\lambda^1:\quad \boldsymbol{W}_{0111}^{0101}\pm\boldsymbol{W}_{0111}^{1101}$$

[4] These may be expressed in matrix form using the sums of four $\boldsymbol{W}$'s.

[5] Note added in proof: In addition to these quadrics the point has also to lie on Q_0+Q_1 (as defined after Theorem 4.2). This does not affect any of the theorems and involves only minor adjustments of the text.

etc., the indices $\begin{smallmatrix}0\\0\end{smallmatrix}$ and $\begin{smallmatrix}1\\0\end{smallmatrix}$ being inserted in the i-th position among $\begin{smallmatrix}101\\111\end{smallmatrix}$ for q_0^i and q_1^i respectively. That is, the nine quadrics Q, q_λ^i define a linear family of freedom 8 depending on the quadrics determining by the nine matrices W_{1111}^{1010}, W_{0111}^{0101}, W_{0111}^{1101}, etc.

These nine do not of course themselves form a Clifford set (since none of them is skew-symmetric), but if each is multiplied by W_{1111}^{0101}, we obtain the Clifford set W_{0000}^{1111}, W_{0001}^{1111}, $W_{0001}^{1110}, \cdots$, which is derived from the primary Clifford set by effecting the interchanges $(1,4)$, $(2,3)$ among the positions of the indices. It follows that $\Phi \equiv fQ + \Sigma f_\lambda{}^i q_\lambda{}^i = 0$ is in general a non-singular quadric, but that, if $(f)^2 + \Sigma f_0{}^i f_1{}^i = 0$, it is a $[7]$-cone.

There are thus ∞^7 $[7]$-cones forming a quadric system in the linear ∞^8 system of quadrics, the vertices of the cones being the invariant spaces, γ, of the singular collineations.

Writing Φ in the form

$$\begin{aligned}\Phi &= x^T(gW_{1111}^{1010} + g_0^1 W_{0111}^{0101} + g_1^1 W_{0111}^{1101} + \cdots)x\\ &= x^T G x, \text{ say,}\end{aligned}$$

it may be verified that

$$x^T \Gamma^T G \Gamma x = x^T G^* x,$$

where the form on the right hand side is that of another quadric of the ∞^8 system. Thus the family of quadrics is invariant under the system of collineations Γ, and the locus $\mathcal{K}$ common to the family is therefore transformed into itself by all collineations Γ. We have then

THEOREM 4.2. *If P lies on $\mathcal{K}$, then it lies in the space, κ, of its conjugates, and κ lies in the meet of the twenty-four primes $\varpi_{\lambda\mu}^{ij}$; κ lies on $\mathcal{K}$.*

In the next section we prove that κ is of dimension 4, and $\mathcal{K}$ of dimension 10.

The nine quadratic forms, Q, $q_\lambda{}^i$, are linearly independent, but are connected by a number of relations quadratic in the coordinates. We may express these most easily by separating the form Q into two parts, thus

$$Q_0 \equiv \Sigma(-1)^\beta x_{\alpha\beta\,\overline{\alpha+\beta}\,0}\, x_{\alpha'\beta'\,\overline{\alpha+\beta}'\,1}$$

$$Q_1 \equiv \Sigma(-1)^\beta x_{\alpha\beta\,\overline{\alpha+\beta}'\,0}\, x_{\alpha'\beta'\,\overline{\alpha+\beta}\,1},$$

so that $Q \equiv Q_0 - Q_1$. Then we have the 16 relations

$$\begin{aligned}x_{\alpha\beta\gamma\delta}Q_\lambda \equiv x_{\alpha'\beta\gamma\delta}q_\alpha{}^1 &+ (-1)^\alpha x_{\alpha\beta'\gamma\delta}q_\beta{}^2\\ &+ (-1)^{\alpha+\beta}x_{\alpha\beta'\gamma\delta}q_\gamma{}^3 + (-1)^{\alpha+\beta+\gamma}x_{\alpha\beta\gamma\delta'}q_\delta{}^4.\end{aligned}$$

where $\lambda' = \alpha + \beta + \gamma + \delta$.

5. The locus of a point which lies in the space of its conjugates. If $W\{^{(\alpha_i)}_{(\beta_j)}\}$ is taken to be a collineation matrix, then from Theorem 1. 2,

$$W\{^{(\alpha_i)}_{(\beta_j)}\}A_{(\gamma_i)} = (-1)^{\Sigma\alpha_i(\beta_i+\gamma_i)}A_{(\beta_i+\gamma_i)}.$$

So that, using the notation of § 3,

$$YA_{\alpha\beta\gamma\delta} = (-1)^{\alpha+\beta+\gamma+\delta}A_{\alpha\beta\gamma\delta}$$

$$Y^1_0 A_{\alpha\beta\gamma\delta} = \tfrac{1}{2}\{(-1)^{\alpha+\beta+\gamma+\delta} + (-1)^{\beta+\gamma+\delta}\}A_{\alpha'\beta\gamma\delta}, \text{ etc}$$

Consider now the spaces of conjugates determined by A_{0000} $(=A)$. Taking $\Gamma_b = bY + \Sigma b^i_0 Y^i_0 + \Sigma b^i_1 Y^i_1$, we have

$$B = \Gamma_b A = bA_{0000} + b^1_0 A_{1000} + b^2_0 A_{0100} + b^3_0 A_{0010} + b^4_0 A_{0001},$$

i. e. the space of conjugates of A_{0000} is the [4],

$\alpha = [A_{0000}, A_{1000}, A_{0100}, A_{0010}, A_{0001}]$, and A and α lie on $\mathcal{K}$. Similarly for any sets of values b, c, d, e we find that $\Gamma_c\Gamma_b A_{0000}$ is a set of points lying entirely in the space determined by all points $A_{\alpha\beta\gamma\delta}$ with two or fewer subscripts "1," and $\Gamma_d\Gamma_c\Gamma_b A$ by all points with three or fewer subscripts "1." That is, the points $\Gamma_c\Gamma_b A$ lie in the [10] α^* given by $x_{1111}, x_{0111}, x_{1011}, x_{1101}, x_{1110} = 0$ and $\Gamma_d\Gamma_c\Gamma_b A$ all lie in the prime, A^*, $x_{1111} = 0$. $\Gamma_e\Gamma_d\Gamma_c\Gamma_b A$ does not in general lie in any element of the simplex $\{A_{\alpha\beta\gamma\delta}\}$.

Any point B of α may be transformed into A_{0000} (by the collineation Γ_b), so that the space β of the conjugates of B is also a [4] lying on $\mathcal{K}$. Next any point C of β may be transformed into B and thence into A_{0000}, so that the space of conjugates of C is a [4], and all these spaces lie on $\mathcal{K}$ and in the prime A^*, forming its complete intersection with $\mathcal{K}$.

Now suppose E is any other point of $\mathcal{K}$. Whatever the coordinates of E, it is clear that at least one of the points YE, $Y^i_\lambda E$ is different from E, so that E has a line of conjugates (at least). This meets A^* in a point D, the space of conjugates of which passes through E. Thus E is also reducible to A, i. e.,

THEOREM 5. 1. *There exist collineations, formed from products of the collineations* Γ, *which transform any point of* $\mathcal{K}$ *into* A_{0000}, *i. e., all points of* $\mathcal{K}$ *are projectively equivalent.*

From this it follows that all properties of $\mathcal{K}$ proved in relation to A_{0000} are valid for all points of $\mathcal{K}$.

$\mathcal{K}$ is the complete intersection of the quadrics Q_λ, q^i_λ, so that its tangent

space at any point is the meet of the tangent primes to the quadrics at the point. From this we find, in particular

THEOREM 5.2. *The* [10] α^* *is the tangent prime to* $\mathcal{K}$ *at* A, *and therefore*

THEOREM 5.3. $\mathcal{K}$ *is of dimension* 10.

THEOREM 5.4. *The tangent-*[10]*'s at all points of* α *lie in the prime* A^*.

The invariant-[7]'s, γ (Theorem 3.6), all lie on Q. Particular among them are the invariant spaces of the collineations $\mathbf{Y}_\lambda^i$. Using the form given at the beginning of this section we find that the invariant space of $\mathbf{Y}_0^1$ is the space determined by the eight points $A_{0\beta\gamma\delta}$. This lies on all quadrics q_λ^i except q_0^1, which it cuts in the non-singular quadric sixfold

$$\bar{q}_0^1 = \Sigma(-1)^\alpha x_{0\alpha\beta 0} x_{0\alpha'\beta' 1} = 0,$$

(with $x_{1\beta\gamma\delta} = 0$).

THEOREM 5.5. *There are* ∞^7 [7]*'s,* γ, *each cutting* $\mathcal{K}$ *in a quadric sixfold.*

I prove next

THEOREM 5.6. *The locus of space* κ_A *which pass through* A *is a cone* $A\mathcal{G}$ *projecting the Grassmannian* $\mathcal{G}_6^5$ *of lines of* [4].

The required locus is the section of $\mathcal{K}$ by α^*; α^* lies on the four quadrics q_1^i, and meets the other quadrics in:

$$\begin{array}{lrrrrrrr}
q_0^1: & & & -x_{0100}x_{0011} & + x_{0010}x_{0101} & - x_{0110}x_{0001} & = 0 \\
q_0^2: & -x_{1000}x_{0011} & & & + x_{0010}x_{1001} & - x_{1010}x_{0001} & = 0 \\
q_0^3: & -x_{1000}x_{0101} & + x_{0100}x_{1001} & & & - x_{1100}x_{0001} & = 0 \\
q_0^4: & -x_{1000}x_{0110} & + x_{0100}x_{1010} & - x_{1100}x_{0010} & & & = 0 \\
Q\ : & -x_{0110}x_{1001} & + x_{1010}x_{0101} & - x_{1100}x_{0011} & & & = 0.
\end{array}$$

In place of $x_{\alpha_1\alpha_2\alpha_3\alpha_4}$, write $y_{ij}, = -y_{ji}$, when $\alpha_i = \alpha_j = 1$ and the other two subscripts are zero, and write $y_{i0}, = y_{0i}$, when $\alpha_i = 1$ and the other three are zero. These equations are then usual equations of the Grassmannian in [9] of lines in [4], except that the signs are written into the equations instead of being carried by the convention $z_{ij} + z_{ji} = 0$.

Properties of $\mathcal{G}$ are most easily discussed in relation to figures in [4]; take then a [4] and name its elements π-points, π-lines, etc. On $\mathcal{G}$ there

is a single ∞^4 system of solids g (each corresponding to π-lines through a π-point), each pair with a single common point (corresponding to the π-line joining two π-points). To the lines in a π-plane there correspond points of a plane s on $\mathcal{B}$: through any line on $\mathcal{B}$ there passes on plane s and one solid g (the line corresponds to a pencil of π-lines of which the vertex is represented by g and the plane by s). Through any point of $\mathcal{B}$ there are ∞^1 solids g and ∞^2 planes s forming a cone of order 3, the complete section by the tangent-[6].[6]

Thus for the cone $A\mathcal{B}$ we find:

THEOREM 5.7. *On $A\mathcal{B}$ in the* [10] α^* *there are* ∞^4 [4]*'s* κ_A *all through* A, *and each pair with a common line, also* ∞^6 [3]*'s* σ, *all through* A. *Through any plane through* A *and on* $A\mathcal{B}$ *there pass one* κ *and one* σ.

A "typical" σ is $[A_{0000}, A_{0100}, A_{0010}, A_{0110}]$.

THEOREM 5.8. *The spaces of conjugates of points of a line through* A *and in* α *all pass through the line and form a cubic line-cone* ($|2, 3|$, [7]).

Associated with $\mathcal{B}$ there are ∞^4 [5]'s (corresponding to π-solids) one through each point of the [9], and each cutting $\mathcal{B}$ in a quadric fourfold, the two systems of generators on which are planes s and section of solids g. Each of these quadrics gives a quadric point-cone in [6] on $A\mathcal{B}$, with generators σ and sections of κ. This point-cone is the section by a tangent prime of the quadric $\bar{q}$ in one of the invariant spaces γ.

In Theorem 5.7 we proved that each [4] κ_A through A meets each other κ_A in a line and no more. Now take B in α, and β the space of its conjugates. α, β meet only in the line AB. Each space κ_B which passes through B meets every other κ_B in a line through B. But amongst the κ_B is α, so that any two [4]'s κ meet in a line (exactly) or not at all, and all the κ's that meet any [4] α of the family in lines lie in the tangent prime A^*.

We may summarize the structure of $\mathcal{K}$ thus:

THEOREM 5.9. *There are on* $\mathcal{K}$: (i) ∞^{10} [4]'s κ *two of which in general do not meet but each meets* ∞^6 *other in lines,* (ii) ∞^{13} [3]*'s* σ, *any pair of which either do not meet, or they may have a point or a line common. Through any plane of* $\mathcal{K}$ *there pass one* κ *and one* σ, (iii) ∞^7 *quadric sixfolds* $\bar{q}$ *on which the generators are solids* σ *and sections of* [4]*'s* κ.

[6] This cone is the "determinantal locus" ($|2, 3|$, [6]). See, e.g., Room (1938), p. 202.

Let us return now to the systems of:

points K, [4]'s κ, [10]'s κ^*, and primes K^* each of freedom 10.

Any set of sixteen homogeneous parameters $\xi_{\alpha\beta\gamma\delta}$, satisfying the ten relations $Q_\lambda = 0$, $q_\lambda^i = 0$, determines: the point K, of which they are the coordinates,

the space κ of its conjugates (from the primes $\varpi_{\lambda\mu}^{ij}$, § 4),

the tangent-[10]'s κ^* to $\mathcal{K}$ at K,

the prime K^* containing the tangent-[10]'s at all points of κ.

Any collineation $\mathbf{\Gamma}$ (or product of $\mathbf{\Gamma}$'s) replaces all these by other spaces of the same system. Any correlation $\mathbf{\Gamma}$ (or product of a correlation $\mathbf{\Gamma}$ with collineations $\mathbf{\Gamma}$) replaces each K by a K^* and each κ by a κ^*.

Thus any incidence relation among a set of points K is reproduced by the corresponding primes K^* and likewise of κ by κ^*. If we can show that there are incidence relations among a set of spaces κ corresponding to given incidence relations among a set of points K, we shall have established a complete *"mutuality"* among all four systems. In general terms the existence of such a correspondence is clear, because each κ is the space of conjugates of a definite point K of itself. The detail of some of these correspondences is:

(i) Two points K_1, K_2 of a space κ determine a line lying in κ; two spaces κ_1, κ_2 which meet in a line determine an ∞^1 system through the line forming (| 2, 3 |, [7]).

(ii) To points of a solid σ correspond the [4]'s κ meeting it in planes.

(iii) On any quadric sixfold $\bar{q}$ there is the Study "triality" linking the points, with the spaces κ the sections of which are one system of generators, and the spaces κ^* the sections of which are the tangent-[6]'s to $\bar{q}$.

We may therefore say

THEOREM 5.10. *There is a "mutuality" among the systems of spaces K, κ, κ^*, K^* analogous to Study's "triality" of points, generator-[3]'s, and tangent primes of a quadric in [7].*

In the general configuration in $[2^r-1]$, we will have a "mutuality" among systems of spaces of dimensions $\nu_0, \nu_1, \nu_2, \cdots, \nu_{r-1}$ where

$$\nu_0 = 0, \nu_1 = r, \cdots, \nu_i = \binom{r}{1} + \binom{r}{2} + \cdots + \binom{r}{i}, \nu_{r-1} = [2^r-2].^7$$

[7] For this sequence of numbers, see for example, Room (1936), p. 539.

Part III.

Two generalizations of the commutative-anti-commutative set.

6. Sets, in general commutative, but with some pairs for which $VU = kUV$. The matrices to be considered have n^r rows and columns. Number the rows and columns in the reversed scale of n, so that $\alpha_1\alpha_2\cdots\alpha_r$ is the index of the row or column with position number $\alpha_1 + n\alpha_2 + \cdots + n^{r-1}\alpha_r$. I make the conventions that:

(i) Greek indices takes the values $0, 1, \cdots, n-1$, and Italic indices take the values $1, \cdots, r$, and

(ii) all arithmetical functions of the Greek indices are reduced modulo n.

Denote the elements of a matrix by $a^{\alpha_1\ldots\alpha_r}_{\beta_1\ldots\beta_r}$, or $a^{(\alpha_i)}_{(\beta_j)}$, and define matrices U_i, V_i thus:

U_i: the diagonal matrix with elements $a^{\alpha_1\ldots\alpha_r}_{\alpha_1\ldots\alpha_r} = \epsilon^{\alpha_i}$, where

$$\epsilon = \exp(2\pi i/n).$$

V_i: the substitution matrix, with one non-zero element in each row and column, given by

$$a^{\alpha_1\alpha_2\ldots\alpha_i\ldots\alpha_r}_{\alpha_1\alpha_2\ldots\alpha_i+1\ldots\alpha_r} = 1,$$

where each subscript except the i-th is equal to the corresponding superscript, while the i-th subscript exceeds by 1 the i-th superscript. For $n = 3$, $r = 2$, the matrices are:

$$U_1 = ((1\epsilon\epsilon^2 1\epsilon\epsilon^2 1\epsilon\epsilon^2)), \quad U_2 = ((111\epsilon\epsilon\epsilon\, \epsilon^2\epsilon^2\epsilon^2)), \text{ where } \epsilon^2 + \epsilon + 1 = 0.$$

V_1, V_2 correspond respectively to the transpositions 231564897, 456789123.

For these matrices we may prove:

THEOREM 6.1. (i) $U_i^n = V_i^n = I$.

(ii) *Every pair of matrices commutes, except the r pairs, U_i, V_i.*

(iii) $V_iU_i = \epsilon U_iV_i$.

THEOREM 6.2. $W^{(\alpha_i)}_{(\beta_j)}, = \Pi U_i^{\alpha_i}\Pi V_j^{\beta_j}$, *is the matrix with one non-zero element in each row and column, given by* $a^{(\lambda_i)}_{(\lambda_j+\beta_j)} = \epsilon^{\Sigma\alpha_i\lambda_i}$.

THEOREM 6.3. $W^{(\alpha_i)}_{(\beta_j)} W^{(\gamma_i)}_{(\delta_j)} = \epsilon^{\Sigma\beta_i\lambda_i} W^{(\alpha_i+\gamma_i)}_{(\beta_j+\delta_j)}$.

THEOREM 6.4. $[W^{(\alpha_i)}_{(\beta_j)}]^n = \begin{cases} I \text{ if } n \text{ is odd,} \\ (-1)\epsilon^{\Sigma\alpha_i\beta_i} I \text{ if } n \text{ is even.} \end{cases}$

THEOREM 6.5. *The* n^{2r+1} *matrices* $\epsilon^{\sigma} W^{(\alpha_i)}_{(\beta_j)}$ *form a group. The* n^{2r} *collineations determined by them form a group.*

THEOREM 6.6. *Any matrix of* n^r *rows and columns may be expressed in a unique way as* $\Sigma k^{(\gamma_i)}_{(\delta_j)} W^{(\gamma_i)}_{(\delta_j)}$, *there being* n^{2r} *terms in the summation.*

Thus, let $\boldsymbol{H} = (h^{(\alpha_i)}_{(\beta_j)})$ be an arbitrary matrix, and let $\tilde{\boldsymbol{H}} = (\tilde{h}^{(\alpha_i)}_{(\beta_j)})$ be the matrix obtained by rearranging the terms in each row in the way determined by

$$h^{(\alpha_i)}_{(\beta_i)} = h^{(\alpha_i)}_{(\alpha_i+\beta_i)}.$$

Then we may prove that

$$\tilde{\boldsymbol{H}} = \boldsymbol{\Omega}\boldsymbol{K}$$

where $\boldsymbol{K} = (k^{(\alpha_i)}_{(\beta_j)})$, and $\boldsymbol{\Omega}$ is the matrix the elements of which are the powers of ϵ given by

$$a^{(\alpha_i)}_{(\beta_i)} = \epsilon^{\Sigma \alpha_i \beta_i}.$$

For example, for $n = 3$, $r = 2$, we have

$$\tilde{\boldsymbol{H}} = \begin{pmatrix} h^{00}_{00} & h^{00}_{10} & h^{00}_{20} & h^{00}_{01} & h^{00}_{11} & h^{00}_{21} & h^{00}_{02} & h^{00}_{12} & h^{00}_{22} \\ h^{10}_{10} & h^{10}_{20} & h^{10}_{00} & h^{10}_{11} & h^{10}_{21} & h^{10}_{01} & h^{10}_{12} & h^{10}_{22} & h^{10}_{02} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ h^{22}_{22} & h^{22}_{02} & h^{22}_{12} & h^{22}_{20} & h^{22}_{00} & h^{22}_{10} & h^{22}_{21} & h^{22}_{01} & h^{22}_{11} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & \epsilon & \epsilon^2 & 1 & \epsilon & \epsilon^2 & 1 & \epsilon & \epsilon^2 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & \epsilon^2 & \epsilon & \epsilon^2 & \epsilon & 1 & \epsilon & 1 & \epsilon^2 \end{pmatrix} \begin{pmatrix} k^{00}_{00} & k^{00}_{10} & k^{00}_{20} & k^{00}_{01} & \cdots & k^{00}_{22} \\ k^{10}_{00} & k^{10}_{10} & k^{10}_{20} & k^{10}_{01} & \cdots & k^{10}_{22} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ k^{22}_{00} & k^{22}_{10} & k^{22}_{20} & k^{22}_{00} & \cdots & k^{22}_{22} \end{pmatrix}.$$

The problems that these matrices suggest immediately are

(i) to pick out Clifford sets of matrices $\boldsymbol{W}$, namely sets no two of whose members commute,

(ii) to pick out sets $\boldsymbol{W}_R$, such that $(\Sigma c_R \boldsymbol{W}_R)^n = f(c_R)\boldsymbol{I}$,

(iii) to find geometrical interpretations of the collineation group determined by the family of matrices.

7. Sets of matrices, in general commutative, but with some pairs for which $\boldsymbol{VU} = \boldsymbol{U}^2\boldsymbol{V}$. Number the rows and columns of a $n^r \times n^r$ matrix as in § 6, and take as basic matrices:

U_i: the diagonal matrix

$$a^{\alpha_1\ldots\alpha_r}_{\alpha_1\ldots\alpha_r} = \eta^{2^{\alpha_i}},$$

where

$$\eta = \exp\{2\pi i/(2^n - 1)\}.$$

V_i: the substitution matrix defined in § 6.

For $n = 3$, $r = 2$, we have:

$$U_1 = ((\eta\ \eta^2\ \eta^4\ \eta\ \eta^2\ \eta^4\ \eta\ \eta^2\ \eta^4)),$$
$$U_2 = ((\eta\ \eta\ \eta\ \eta^2\ \eta^2\ \eta^2\ \eta^4\ \eta^4\ \eta^4)),$$

where $\eta^7 = 1$.

Then:

THEOREM 7.1. (i) $U_i^{2^n-1} = V_i^n = I$.

(ii) *Every pair of matrices commutes, except the r pairs U_i, V_i.*

(iii) $V_iU_i = U_i^2V_i$.

THEOREM 7.2. $V_i^\beta U_i^\alpha = U_i^{2^\beta\alpha}V_i^\beta$.

THEOREM 7.3. *If* $W^{(\alpha_i)}_{(\beta_j)} = \Pi U_i^{\alpha_i}\Pi V_j^{\beta_j}$, *then*

$$W^{(\alpha_i)}_{(\beta_j)} W^{(\gamma_i)}_{(\delta_j)} = W^{(\alpha_i+2^{\beta_i}\gamma_i)}_{(\beta_j+\delta_j)},$$

where the superscripts of W are reduced modulo $2^n - 1$, and the subscripts modulo n.

THEOREM 7.4. *The matrices W form a group in which*

$$[W^{(\alpha_i)}_{(\beta_i)}]^{-1} = W^{(-2^{n-\beta_i}\alpha_i)}_{(n-\beta_i)}.$$

8. Further generaliaztions. Let $n_1, n_2, \cdots, n_r$ be any r positive integers, not necessarily all different, each greater than 1, and let $N = \Pi n_i$, $i = 1, \cdots, r$. Write the position numbers, beginning with 0, of the rows and columns of a N-by-N matrix in the form

$$\alpha_1 + n_1\alpha_2 + n_1n_2\alpha_3 + \cdots + n_1n_2\cdots n_{r-1}\alpha_r,$$

where α_i is to take the values $0, 1, \cdots, n_i - 1$. Take $a^{\alpha_1\ldots\alpha_r}_{\beta_1\ldots\beta_r}$ or $a^{(\alpha_i)}_{(\beta_j)}$ as the elements of a matrix. Then construct the two sets of r matrices:

U_i: the diagonal matrix in which

$$a^{\alpha_1\ldots\alpha_r}_{\alpha_1\ldots\alpha_r} = \epsilon^{\alpha_i}, \text{ where } \epsilon_i = \exp(2i\pi/n_i)$$

V_i: the substitution matrix with non-zero elements $a^{\alpha_1\ldots\alpha_i\ldots\alpha_r}_{\alpha_1\ldots\alpha_i+1\ldots\alpha_r} = 1$, each subscript except the i-th being equal to the corresponding superscript. E.g., for $n_1 = 2$, $n_2 = 3$, $n_3 = 2$ the matrices are

$$U_1 = ((1\quad -1\quad 1\quad -1\quad 1\quad -1\quad 1\quad -1\quad 1\quad -1\quad 1\quad -1)),$$

$$U_2 = ((1\quad 1\quad \epsilon\quad \epsilon\quad \epsilon^2\quad \epsilon^2\quad 1\quad 1\quad \epsilon\quad \epsilon\quad \epsilon^2\quad \epsilon^2)),$$
$$\text{where } \epsilon^2 + \epsilon + 1 = 0,$$

$$U_3 = ((1\quad 1\quad 1\quad 1\quad 1\quad 1\quad -1\quad -1\quad -1\quad -1\quad -1\quad -1)).$$

V_1, V_2, V_3 correspond to the substitutions: $1032547698et$, $23450189te67$, $6789te012345$.

These matrices have properties analogous to those discussed in § 6, the difference being that the number n_i involved is different for each pair of matrices U_i, V_i, instead of being the same for all pairs.

THEOREM 8.1. (i) $U_i^{n_i} = I$, $V_i^{n_i} = I$.

(ii) *Every pair of matrices commutes except the r pairs* U_i, V_i.

(iii) $V_iU_i = \epsilon_i U_i V_i$, *where* $\epsilon_i = \exp(2i\pi/n_i)$.

THEOREM 8.2. *If* $W^{\alpha_1\cdots\alpha_r}_{\beta_1\cdots\beta_r} = \Pi U_i^{\alpha_i}\Pi V_i^{\beta_i}$, *then*

$$W^{(\alpha_i)}_{(\beta_j)}\, W^{(\gamma_i)}_{(\delta_j)} = (\Pi\epsilon_i^{\beta_i\gamma_i})\, W^{(\alpha_i+\gamma_i)}_{(\beta_j+\delta_j)},$$

where the i-th index is reduced modulo n_i.

THEOREM 8.3. *Any N-by-N matrix may be expressed in a single way in the form* $\Sigma k^{(\alpha_i)}_{(\beta_j)}\, W^{(\alpha_i)}_{(\beta_j)}$ *for a given ordered set of factors* $n_1, n_2, \cdots, n_r$ *of N.* Thus, if $H = (h^{(\alpha_i)}_{(\beta_i)})$ is the matrix, then, writing

$$\tilde{H} = (\tilde{h}^{(\alpha_i)}_{(\beta_i)}) \text{ where } \tilde{h}^{(\alpha_i)}_{(\beta_i)} = h^{(\alpha_i)}_{(\alpha_i+\beta_i)}, \qquad K = (k^{(\alpha_i)}_{(\beta_i)}),$$

$$\Omega = \text{the matrix with elements } a^{\alpha_1\alpha_2\cdots\alpha_r}_{\beta_1\beta_2\cdots\beta_r} = \Pi\epsilon_i^{\alpha_i\beta_i},$$

we have $\tilde{H} = \Omega K$.

E. g., for $n_1 = 3$, $n_2 = 2$:

$$\begin{bmatrix} h^{00}_{00} & h^{00}_{10} & h^{00}_{20} & h^{00}_{01} & h^{00}_{11} & h^{00}_{21} \\ h^{10}_{10} & h^{10}_{20} & h^{10}_{00} & h^{10}_{11} & h^{10}_{21} & h^{10}_{01} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ h^{21}_{21} & h^{21}_{01} & h^{21}_{11} & h^{21}_{20} & h^{21}_{00} & h^{21}_{10} \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & \epsilon & \epsilon^2 & 1 & \epsilon & \epsilon^2 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & \epsilon^2 & \epsilon & -1 & -\epsilon^2 & -\epsilon \end{bmatrix} \begin{bmatrix} k^{00}_{00} & k^{00}_{10} & \cdots & k^{00}_{21} \\ k^{10}_{00} & k^{10}_{10} & \cdots & k^{10}_{21} \\ \cdot & \cdot & \cdot & \cdot \\ k^{21}_{00} & k^{21}_{10} & \cdots & k^{21}_{21} \end{bmatrix}.$$

The family of matrices discussed in § 7 may be generalized in the same way, and may be generalized also by defining the η_i in terms of powers which are powers of numbers t_i different from 2, so as to obtain relations $V_iU_i = U_i^{t_i}V_i$.

UNIVERSITY OF SYDNEY (AUSTRALIA).

REFERENCES.

[1] E. Cartan, "La Théorie des Spineurs, II," *Actualités scientifiques et industrielles*, no. 701 (1938).

[2] N. H. Kuiper, *American Journal of Mathematics*, vol. 72 (1949), pp. 425-441.

[3] T. G. Room, *Proceedings of the London Mathematical Society* (2), vol. 42 (1937), pp. 529-545.

[4] ———, *Geometry of Determinantal Loci*, Cambridge (1938).